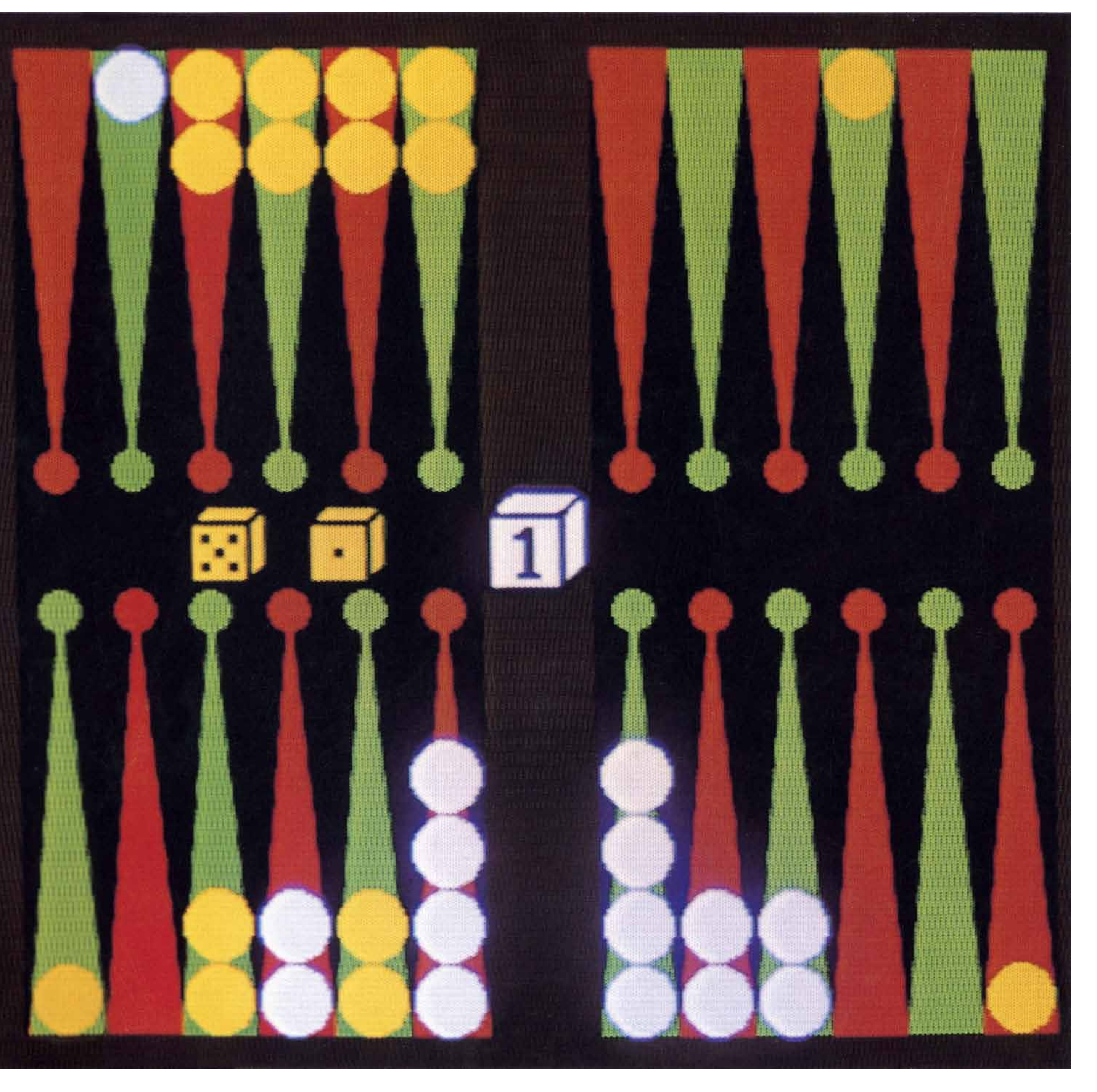
INVESTIGACION Y CIENCIA

Edición en español de SCIENTIFIC AMERICAN



CHAQUETE

Agosto 1980

250 Ptas.

Los espacios en gris
corresponden a publicidad
en la edición impresa

Número 47
Agosto 1980

INVESTIGACION Y
CIENCIA
Edición en español de SCIENTIFIC AMERICAN

EDITA Prensa Científica, S.A.
Calabria, 235-239
Barcelona-29 (ESPAÑA)

Colaboradores de este número:

Asesoramiento y traducción:

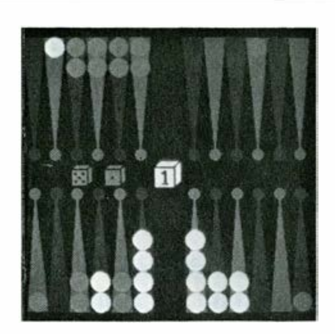
Manuel Puigcerver: *Meteoritos basálticos;* Luis Bou: *Chaquete por ordenador;* Cristóbal Mezquita: *Fisiología de las plaquetas sanguíneas;* Pedro Pascual: *Teorías gauge de las fuerzas entre partículas elementales;* Jesús A. Ortea y Germán Flor: *Subsistencia en el norte de España durante la última glaciación;* Montserrat Domingo: *La hipótesis del gas de las profundidades terrestres;* Joandomènec Ros: *Selección sexual en* Hylobittacus apicalis; Luis Bou: *Juegos matemáticos;* J. Vilardell: *Taller y laboratorio.*

Ciencia y sociedad:

Fritz Fuchs

Libros:

Domingo Fletcher, M.ª Asunción Catalá y Luis Alonso



LA PORTADA

La fotografía de portada es una visualización producida mediante ordenador de una situación crítica de la quinta y última partida del encuentro celebrado entre el campeón del mundo de chaquete, Luigi Villa, y un programa de ordenador diseñado para este juego, BKG 9.8. El programa, que consiguió vencer a Villa por el tanteo de 7 a 1, se ha convertido así en el primer programa de ordenador que haya derrotado a un campeón mundial, ya sea de juegos de cartas o de juegos de tablero. Villa disponía de las piezas blancas, que avanzan en el sentido horario; BKG 9.8 conducía las anaranjadas, que se desplazan en sentido antihorario. En esta situación, BKG 9.8, cuyos dados acababan de darle un cinco y un uno, ejecutó una jugada sensacional, que sorprendió a sus programadores y mereció los aplausos de los expertos que presenciaban la partida (véase "Chaquete por ordenador", por Hans Berliner, en este mismo número). La visualización de ordenador ha sido proporcionada por la Three Rivers Computer.

Suscripciones:
Prensa Científica, S. A.
Calabria, 235-239
Barcelona-29 (España)
Teléfono 322 05 51 ext. 41

Condiciones de suscripción:
España:
Un año (12 números): 2.750 pesetas
Extranjero:
Un año (12 números): 43 U.S.$
Ejemplar atrasado ordinario:
280 pesetas
Ejemplar atrasado extraordinario:
420 pesetas

Distribución para España:
Distribuciones de Enlace, S.A.
Ausias March, 49, Barcelona-10

Distribución para los restantes paises:
Editorial Labor, S. A.
Calabria, 235-239, Barcelona-29

Publicidad:
Madrid:
Gustavo Martínez Ovin
Avda. de Moratalaz, 137, Madrid-30
Tel. 430 84 81
Cataluña:
Miguel Munill
Balmes, 191, 2.º, 2.ª, Barcelona-6
Tels. 218 44 45 y 218 40 86

Controlado
por O.J.D.



PROCEDENCIA DE LAS ILUSTRACIONES

La portada es cortesía de Ralph Morse

Página	Fuente
9-10	Miguel Alonso
11-13	Concepción Azcón - G. de Aguilar y José Miguel Barea
14-15	Miguel Alonso
16	Concepción Azcón - G. de Aguilar y José Miguel Barea
19	Dane A. Penland
20	Harry Y. McSween, Jr., Universidad de Tennessee
21	Walken Graphics
22	National Aeronautics and Space Administration
23	Walken Graphics
24	Walken Graphics (*arriba y abajo izquierda*) Harry Y. McSween, Jr., (*abajo derecha*)
25-26	Walken Graphics
27	Jet Propulsion Laboratory
31	Ralph Morse
32-39	Jerome Kuhl
41	Carol Donner
42	División de Ultrasonidos New York Hospital-Cornell Medical Center
47	Hans R. Baumgartner, F. Hoffmann-La Roche & Co.
48	James G. White, Universidad de Minnesota (*izquierda*); Patricia J. Wynne (*derecha*)
49	Patricia J. Wynne
50	James G. White, Universidad de Minnesota
51-54	Patricia J. Wynne
55	Hans R. Baumgartner, F. Hoffmann-La Roche & Co.
56-57	Hans R. Baumgartner, F. Hoffmann-La Roche & Co. (*arriba*) Patricia J. Wynne (*abajo*)
59-75	Allen Beechel
79	Lawrence Guy Straus, Universidad de Nuevo México
80-81	Andrew Tomko
82-86	Alan D. Iselin
89-93	George V. Kelvin
94-95	Sovfoto
99	Randy Thornhill, Universidad de Nuevo México
100-104	Tom Prentiss
108-112	Ilil Arbel
115	Alan L. Fald (*arriba*) Jearl Walker (*abajo*)
116-120	Michael Goodman

ISSN 0210-136X
Dep. legal: B. 38.999-76
Fotocomposición Tecfa
Guipúzcoa, 36 (local 1) Barcelona-20
Imprime Grafesa. Nápoles, 249
Barcelona-13
Printed in Spain - Impreso en España

Los autores

CONCEPCION AZCON GONZALEZ DE AGUILAR y JOSE MIGUEL BAREA ("Micorrizas") desarrollan su labor de investigación en el departamento de microbiología de la Estación Experimental del Zaidín (Granada) del Consejo Superior de Investigaciones Científicas (CSIC). Han centrado su campo de estudio en la ecología y mecanismos bioquímicos de acción de los microorganismos del suelo y su influencia en nutrición vegetal. Barea, doctor en farmacia por la Universidad de Granada, es investigador científico del CSIC, organismo en el que trabaja desde 1966, si bien, en los primeros años de su carrera científica, simultaneó esta función con la docente en el departamento de microbiología de la Universidad de Granada. Trabajó en la Estación Experimental de Rothamsted (Inglaterra) en 1972, especializándose en diversos aspectos de la microbiología del suelo, fundamentalmente en el estudio de las micorrizas vesículo-arbusculares. Concepción Azcón es doctora en ciencias biológicas y concentra su labor científica en el estudio de las simbiosis microbio-planta, dada la repercusión de las mismas en el crecimiento y nutrición de los vegetales. Acaban de publicar "Effects of microorganisms in the nutritional requeriments of plants" (1980).

HARRY Y. McSWEEN, Jr., y EDWARD M. STOLPER ("Meteoritos basálticos") comenzaron su colaboración siendo estudiantes de segundo ciclo en la Universidad de Harvard. McSween, que es ahora profesor adjunto de geología en la Universidad de Tennessee, siguió su primer ciclo en química en The Citadel y recibió su licenciatura (en geología) por la Universidad de Georgia. Tras varios años en las Fuerzas Aéreas como piloto, reanudó sus estudios superiores, obteniendo el doctorado en Harvard en 1977. Stolper se graduó por el Harvard College en 1974. Después de dos años como becario Marshall en la Universidad de Edimburgo, donde recibió su licenciatura en geología, regresó a Harvard y se doctoró en 1979.

HANS BERLINER ("Chaquete por ordenador") es investigador "senior" en ciencias de cómputo, en la Carnegie-Mellon University. Detenta también el título de gran maestro de ajedrez por correspondencia. Tras graduarse en psicología por la George Washington University, en 1954, trabajó en el campo de factores humanos durante siete años, hasta que optó por familiarizarse más con los ordenadores. Desde 1961 hasta 1969 trabajó en la división de sistemas para la administración federal de la IBM, contribuyendo a desarrollar grandes sistemas de proceso de datos para el gobierno de Estados Unidos. En 1969 decidió centrar sus esfuerzos en temas de investigación relacionados con ordenadores, sobre todo, los relativos a inteligencia artificial, y asistió a la Carnegie-Mellon University, donde alcanzó el grado de doctor en ciencias de cómputo en el año 1975.

MARJORIE B. ZUCKER ("Fisiología de las plaquetas") es profesora de patología de la Facultad de Medicina de la Universidad de Nueva York. Se graduó en el Vassar College en 1939 y obtuvo el doctorado en fisiología por la Universidad de Columbia en 1944. "Enseñé fisiología inicialmente", escribe, "y puse en marcha las pruebas de coagulación en el Memorial Hospital-Sloan-Kettering Institute. En 1962, empecé a trabajar en los laboratorios de investigación de la Cruz Roja Americana en el hospital de la Universidad de Nueva York. Cuando la Cruz Roja estableció su centro de investigación en Bethesada, permanecí en la Universidad de Nueva York". Zucker es coautora con Aaron Marcus de un libro sobre plaquetas.

GERARD 'T HOOFT ("Teorías gauge de las fuerzas entre partículas elementales") enseña física en la Universidad de Utrecht, donde se graduó en 1969 y se doctoró en 1972. Allí volvió como docente tras permanecer dos años en la Organización Europa para Investigaciones Nucleares (CERN). "Mi interés por la física se remonta a los seis años de edad. Dudé entre la pintura, la música y la física. Opté por la física, inspirado por otros físicos de mi familia, pero aún me gusta pintar un cuadro y tocar el piano de vez en cuando."

LAWRENCE GUY STRAUS, GEOFFREY A. CLARK, JESUS ALTUNA y JESUS A. ORTEA ("Subsistencia en el norte de España durante la última glaciación") trabajan en equipo en el campo de la antropología y arqueología española, bajo los auspicios de la National Sciencie Foundation. Straus, que se formó en Chicago, da clases de antropología en la Universidad de Nuevo México. Clark enseña la misma disciplina en la Universidad de Arizona, donde realizó sus estudios. Altuna dirige el laboratorio de paleontología de la Sociedad de Ciencias Naturales Aranzadi, de San Sebastián, y obtuvo su doctorado en ciencias biológicas por la Universidad de Madrid. Ortea, profesor de zoología de la Universidad de Oviedo, recibió en ese centro superior el título de doctor en biología. Dirige el Servicio de Investigaciones Marinas de la Consejería de Pesca de Asturias.

THOMAS GOLD y STEVEN SOTER ("La hipótesis del gas de las profundidades terrestres") disfrutan de un permiso de la Universidad de Cornell. Gold, que es profesor de astronomía del Centro de Investigación Radiofísica y Espacial, se encuentra en el Instituto Niels Bohr de Copenhague. Soter, investigador asociado del centro, se halla en la estación de televisión KCET de Los Angeles realizando la producción "Cosmos", un documental sobre astronomía. Gold nació en Viena y obtuvo su titulación en la Universidad de Cambridge: su bachillerato y licenciatura, ambos en ciencias mecánicas, en 1942 y 1946, respectivamente, y su doctorado en 1969. Trabajó en Inglaterra y luego en Cambridge (Massachusetts), hasta 1957, año en que pasó a ser profesor de astronomía en la Universidad de Harvard; de allí se trasladó a Cornell, en 1959, donde dirige el departamento de astronomía. Soter se graduó por la Universidad de California en Los Angeles en 1965; obtuvo su doctorado en astronomía por la de Cornell en 1971.

RANDY THORNHILL ("Selección sexual en *Hylobittacus apicalis*") es profesor ayudante de biología en la Universidad de Nuevo México. "Obtuve mi licenciatura en la Auburn University", escribe, "y mi doctorado (en zoología) por la Universidad de Michigan en 1974. Estuve en la facultad de la Universidad de Florida antes de trasladarme a Nuevo México en 1975. En mi tiempo libre gusto de realizar actividades con mi familia, en particular acampar, que combino con la recolección y observación de insectos".

Hace...

José M.ª López Piñero

...trescientos años

Crisóstomo Martínez comenzó a trabajar en su atlas anatómico. Nacido en Valencia el año 1638, había desplegado hasta entonces una notable actividad como grabador y también como pintor. Están identificados como suyos unos quince grabados en metal y algunos en marfil, aparte de varios dibujos y de los cuadros que sabemos que pintó y que en su casi totalidad se han perdido.

En noviembre de 1685, las autoridades de la ciudad de Valencia y los catedráticos de medicina de su universidad solicitaron a Carlos II permiso para concederle una importante ayuda económica. En diciembre del año siguiente el monarca concedió tal permiso. Gracias a varias cartas que Martínez escribió a Juan Bautista Gil de Castelldases, catedrático de medicina de la Universidad de Valencia, sabemos que llegó a París el 19 de julio de 1687, consagrándose allí a terminar el trabajo iniciado en Valencia. Residió en el famoso Collège de Montaigue y se puso en relación con el anatomista Guichard Joseph du Verney y con el ambiente científico de la Académie des Sciences, entonces el más avanzado de París. La última carta, escrita a finales de 1689, refleja las duras condiciones de su estancia en la capital francesa, a causa de la gota que padecía y de las dificultades que le produjo la guerra entre Francia y los componentes de la Liga de Augsburgo, de la que formaba parte España. En 1690, tuvo que abandonar París, al ser acusado de espía. No disponemos de ninguna noticia directa de lo que le sucedió con posterioridad a esta fecha. Un contemporáneo y compatriota suyo, el P. José Rodríguez, afirma que murió en Flandes en 1694. En cualquier caso, su atlas no llegó jamás a ser editado.

En el Archivo Histórico Municipal de Valencia, junto a las cartas citadas, se conservan dieciocho láminas anatómicas originales de Crisóstomo Martínez y siete manuscritos suyos que contienen la explicación de seis de ellas y un texto titulado «Generalidades acerca de los huesos». En vida de su autor, solamente fue editada una de estas láminas, que es un grabado de gran tamaño (67,7 × 50,8 cm.) dedicado al estudio de las proporciones en tres figuras humanas denudadas y un esqueleto infantil. Apareció en París en 1689 y tres años más tarde fue reimpresa en Frankfurt y Leipzig.

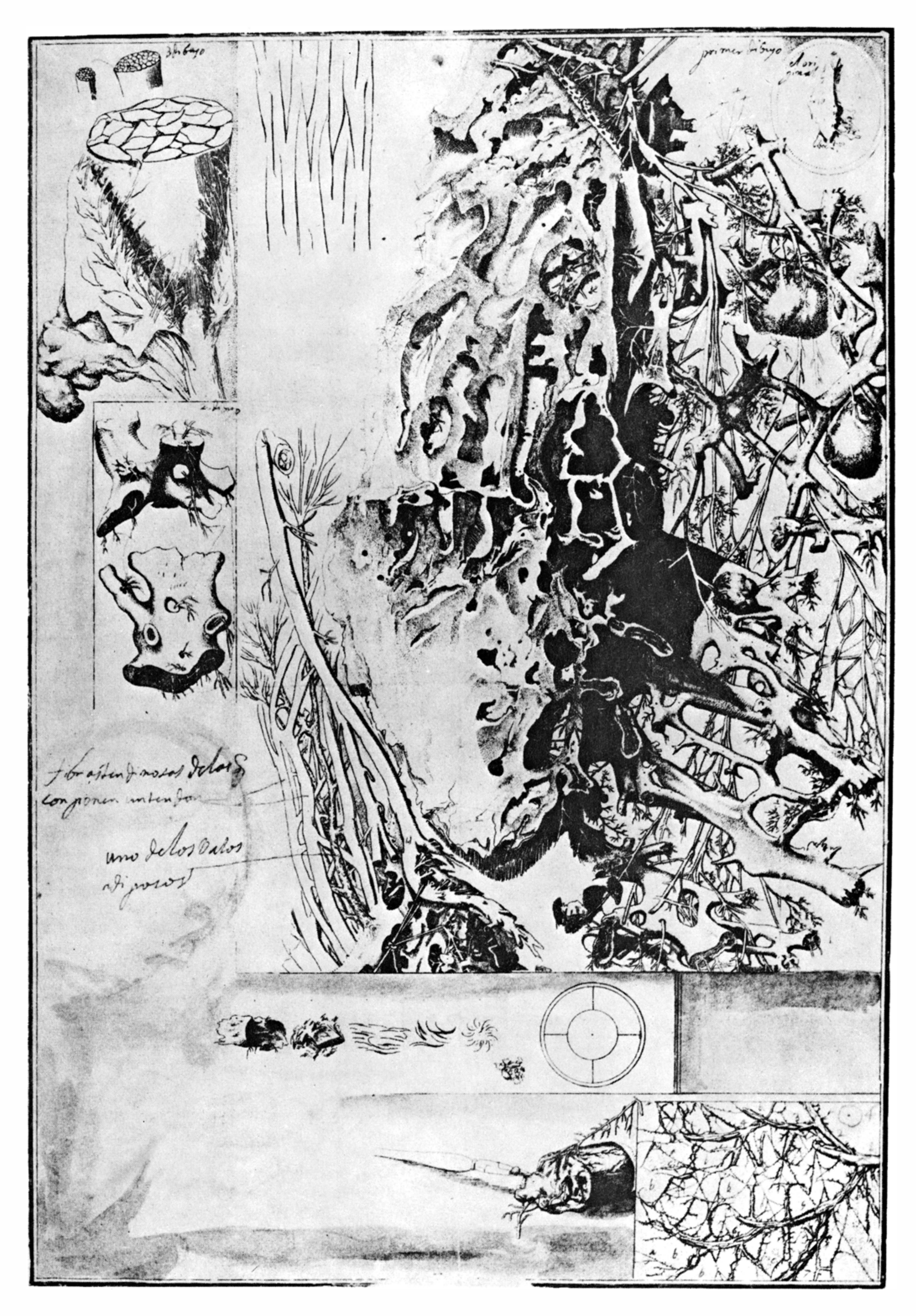
La obra de Crisóstomo Martínez debió alcanzar gran prestigio en el París de la época porque medio siglo después, en 1740, se publicó en dicha ciudad una reedición de la misma lámina, acompañada esta vez de otra de parecido tamaño que habría grabado el artista valenciano. Esta última representa en su parte superior a doce esqueletos en distintas actitudes, y en la inferior a diferentes huesos enteros o cortados longitudinalmente, junto a varios esquemas y algunas ampliaciones microscópicas y un cráneo infantil visto al trasluz. Acompaña a las dos láminas un folleto que incluye un «Éloge» de Crisóstomo Martínez redactado por el anónimo editor, el texto explicativo de la lámina de las proporciones que el propio grabador valenciano había publicado en 1689, y otra explicación de la lámina osteológica escrita también por el editor, aunque revisada por el gran anatomista Jacobus Benignus Winslow. Las dos láminas y sus explicaciones fueron muy estimadas en la enseñanza de la anatomía para los artistas. La Académie Royale de Peinture adquirió las planchas y publicó en 1780 una reimpresión de ambos grabados y del folleto con idéntico texto que apareciera en la edición de cincuenta años antes.

El resto de la producción de Crisóstomo Martínez quedó inédito. Ello no quita a las láminas y manuscritos conservados en Valencia su extraordinario interés como fuente de una de las más importantes contribuciones del movimiento científico renovador que se desarrolló en España a finales del siglo diecisiete.

Las láminas contienen representaciones macroscópicas de la mayor parte de los huesos y músculos del organismo humano que, por su calidad, constituyen un aportación de excepcional importancia dentro de la historia de la iconografía anatómica en España. Resulta patente que tanto los grabados como sus explicaciones responden a un vivo interés por la interpretación funcional de las estructuras anatómicas, característica típica, como es sabido, de la morfología del Seiscientos. Propia también de la época es la preocupación de Martínez por la embriología ósea, que le lleva a exponer en uno de sus manuscritos una doctrina embriológica general basada en el preformacionismo, es decir, la teoría «moderna» que en el siglo XVII se opuso a las clásicas doctrinas aristotélicas.

La parte más nueva e importante de la obra de Crisóstomo Martínez fue, sin embargo, la dedicada a investigar la fina estructura ósea por distintos medios y muy en primer término con el microscopio. El grabador y anatomista valenciano perteneció, en efecto, con pleno derecho a la primera generación de microscopistas europeos. Coetáneo de Marcello Malpighi, Antonj van Leeuwenhoek y Robert Hooke, sus grabados y sus escritos corresponden, no al hombre ocasionalmente interesado por el nuevo medio de observación, sino al investigador con hábito de trabajo que ha penetrado en las dificultades técnicas y en las precauciones que hay que tomar para asegurar la objetividad de lo observado. «Para ver y examinar bien esto —dice, hablando del estudio microscópico de los huesos— se rompe de alto abajo un pedazo de hueso (yo digo hueso suponiendo que las epífisas, lo están ya, como la parte principal y sus asperezas, que tenga adherencia con él, una parte de un tendón), y de esta suerte se ve lo interno, lo externo y lo intrínseco, y rompiendo de dicho hueso algunos pedacillos en diversos modos y algunos de ellos tenuísimos, quedando siempre adherentes al dicho tendón, se averigua en este modo esta verdad.

Lámina XI del mismo atlas. Varios dibujos aislados con preparaciones microscópicas de un fragmento óseo cuyo tamaño original se representa en el ángulo superior derecho. El tema principal es la irrigación vascular y la fina estructura de la médula ósea. No está acabada.


primer dibuxo
fibras tendinosas de los
componen un tendon
uno de los vasos
adiposos

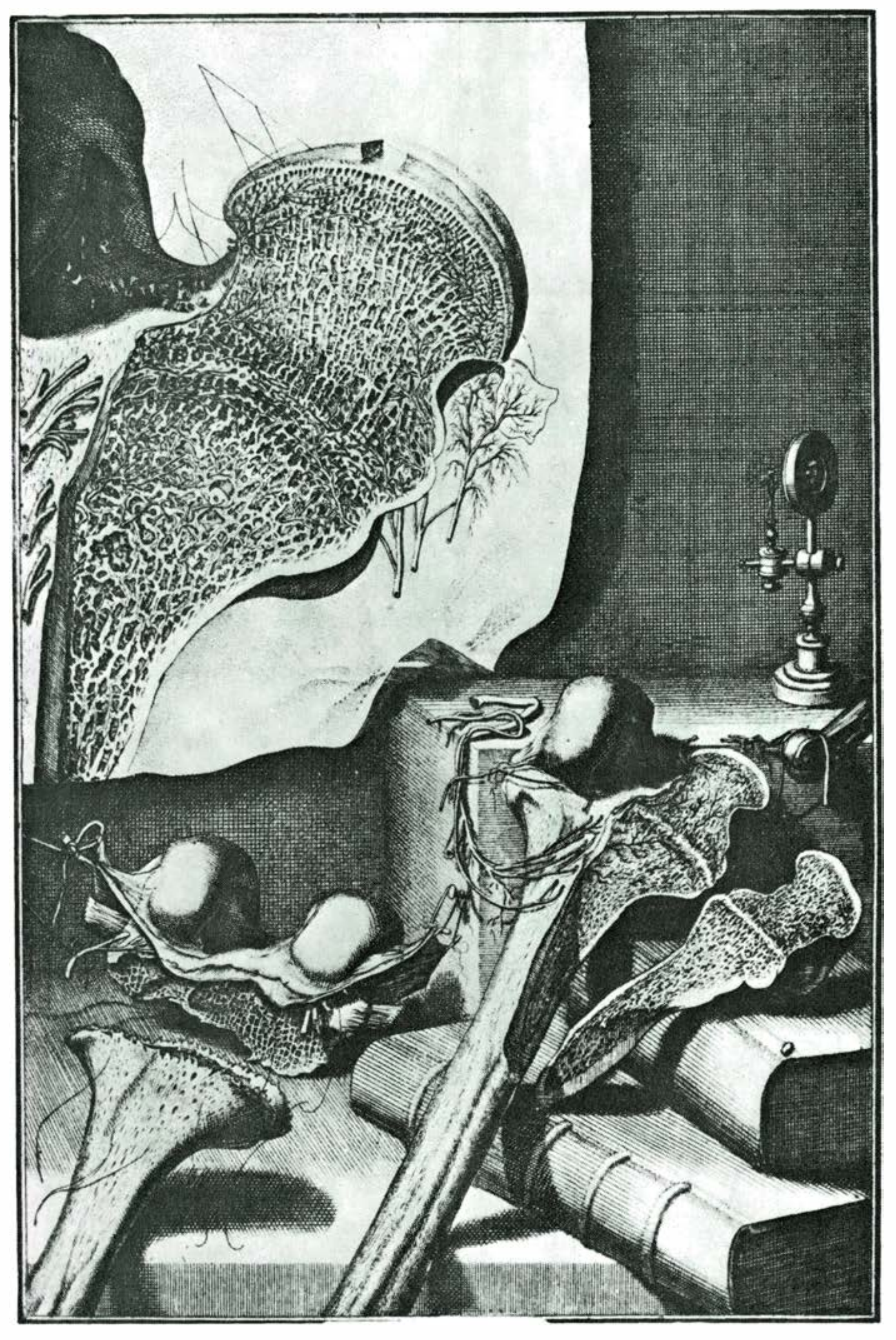


Lámina V del atlas de Crisóstomo Martínez. Cortes transversales y longitudinales de la cabeza y del cuello del fémur izquierdo. Ampliación que muestra la textura del tejido esponjoso y la penetración y ramificación de los vasos nutricios. Microscopio.

Esto requiere mucha maña y diversidad de huesos, unos crudos, otros cocidos y otros secos o medio secos, y variedad de vidrios, esto es, unos que descubran una gran parte, con aumento y claridad fiel, para hacerse capaz de lo total; después se examina una parte de esta parte, con otro vidrio que aumente más, y así, por grados, hasta llegar a examinar con un microscopio muy fino una partecilla tenuísima, y como se ha venido a la especulación de esta partecilla desde una muy grande por una segura gradación, queda con certidumbre averiguada esta mínima y tenuísima porción, corriendo seguro por estos grados, desde lo grande a lo pequeño y de lo pequeño a lo grande; y no haciéndolo de esta suerte se corre agradablemente el engaño, porque es verdad que un microscopio descubre admirablemente las cosas, con todo eso, si se consulta sin más ni más, tal vez desfigura las cosas, esto es, como todo lo aumenta, aumenta también la luz de que los objetos están húmedos, mojados u oleaginosos, y añadiendo a esta luz la que los vidrios reflejan, resulta que una simple membrana parece una tela de plata con maravillosa labor que alegra y deja absorta la imaginación, y no sólo plieguecito en ella parece un nervio que se va ramificando, unas desigualdades, asperezas o eminencias tocadas vivamente de luz parecen verdaderamente unas vejiguillas llenas de algún humor, y a este modo se ofrecen muchos engaños, y así es menester mucho examen, mucha cautela y prudencia y tiempo, para no engañarse».

El objeto de las investigaciones microscópicas de Martínez fue también rigurosamente moderno. No siguió trabajando con el hueso desnudo y seco, como la estática anatomía renacentista. Con una intención al mismo tiempo estructuralista y funcional, se centró en la llamada «osteología fresca», indagando la inserción de los ligamentos, el acoplamiento de los cartílagos, la irrigación e inervación de los huesos, la estructura y el significado funcional de la médula ósea, etcétera.

La irrigación ósea, en concreto, fue el tema al que mayores esfuerzos dedicó. Su lámina XI, la más interesante desde el punto de vista microscópico, y su texto explicativo están consagrados al análisis de la distribución y conexiones de la fina irrigación arterial y venosa de los huesos y de su relación con la inervación. En este contexto defiende la existencia de unas vesículas microscópicas situadas en la extremidad de las más finas ramificaciones vasculares, y destinadas a recibir la grasa, que proviene de la sangre por filtración. El conjunto de tales vesículas forma el tuétano o médula ósea. A partir de aquí, la grasa se distribuye a través de un sistema de vasos, distinto de las arterias, venas, linfáticos y nervios: los «vasos adiposos», cuya distribución llegó incluso a investigar en la fibra muscular.

Esta teoría de Crisóstomo Martínez —una de las escasas contribuciones originales de un español a la ciencia de la época— refleja muy expresivamente el carácter «moderno» del pensamiento de su autor. Su interpretación de la médula ósea se oponía rotundamente a la del galenismo tradicional, que la consideraba medio nutricio del hueso. Los fundamentos inmediatos de sus hipótesis fueron, por el contrario, los esquemas iatroquímicos y la doctrina de la circulación de la sangre.

Micorrizas

Estas simbiosis hongo-planta desempeñan un importante papel en ecología y nutrición vegetal. Formadas por la mayoría de las plantas existentes, comienza a estudiarse su aplicabilidad en agricultura como fertilizante microbiano

Concepción Azcón-G. de Aguilar y José Miguel Barea

Hace unos 400 millones de años las plantas comenzaron a colonizar la superficie terrestre, hecho éste importante en la evolución de los seres vivos. Mediante su capacidad fotosintética, los vegetales transforman la energía solar en energía química, utilizable por otros organismos. Para llevar a cabo tal actividad, aquellas primeras plantas requerirían, como las actuales, una serie de nutrientes que deberían captar de la atmósfera (anhídrido carbónico) y del suelo (agua, nitrógeno, fósforo, azufre, potasio, manganeso, hierro, etcétera).

Mayoritariamente se acepta que las plantas se originaron a partir de las algas verdes. El tránsito de éstas, desde su hábitat acuático a ambientes secos, y la evolución en ellos a plantas con raíces, antecesoras de los actuales vegetales superiores, fue en efecto punto crucial en la historia de la vida sobre la tierra. Sin lugar a dudas, el hambre y la sed, las dos grandes y eternas dificultades de la existencia sobre nuestro planeta, incidieron de forma decisiva en los primeros pasos de la evolución de los vegetales. La opinión común sostiene que la colonización de aquel suelo, seco y pobre, por las algas fue posible gracias a que éstas se asociaron con microorganismos, lo cual permitió que pudieran captar sus alimentos minerales. De un lado, hongos microscópicos formaron las primeras micorrizas, simbiosis especializadas en la captación de fósforo, y de otro, las "plantas" se asociaron con microorganismos fijadores de nitrógeno molecular (N_2) atmosférico.

Las micorrizas son simbiosis mutualísticas entre hongos y raíces de plantas superiores. Salvo en contadas excepciones, la planta suministra al hongo fuentes de carbono procedentes del producto de la fotosíntesis, además de un nicho ecológico protegido de los fenómenos de antagonismo microbiano en la rizosfera. Por su parte, el hongo ayuda a la planta a absorber sus nutrientes minerales del suelo. Se sabe que las hifas del hongo que se desarrollan en la raíz y emergen de ella desempeñan un importante papel en la translocación hacia la planta de iones fosfato, por lo que, en suelos con un contenido bajo en fósforo asimilable, caso generalizado en la mayoría de los suelos agrícolas, las micorrizas representan una contribución fundamental para la economía nutritiva de la planta. (En los hongos, se llama hifa a cada uno de los elementos filamentosos que constituyen su aparato vegetativo, el micelio.)

Debido a la amplia distribución de las micorrizas puede afirmarse que las plantas cuando crecen en condiciones naturales son en su mayoría organismos dobles, en el sentido de que el órgano a través del cual absorben agua y nutrientes está constituido por la raíz propiamente dicha y un hongo que vive simbióticamente con ella.

Las micorrizas se han venido clasificando, en base a su estructura y morfología, en dos grandes grupos: ectotróficas y endotróficas. En las primeras se incluían micorrizas en las cuales el hongo, normalmente de micelio tabicado, forma un auténtico manto de hifas que rodea la raíz. El desarrollo del hongo en el interior de la corteza es intercelular, dando aspecto de red (red de Hartig). En las endotróficas, sin embargo, el hongo no forma manto sobre la raíz, y las hifas penetran en el interior de las células de la corteza. No obstante, hoy día se sabe que los hongos formadores de endomicorrizas están muy distanciados taxonómica y fisiológicamente, por lo que ha sido necesario modificar esta clasificación y subdividir a las antiguas micorrizas endotróficas en varios grupos. Sin lugar a dudas, las más extendidas son las de tipo vesículo-arbuscular (VA), ya que esta simbiosis se encuentra en todos los climas que permiten el desarrollo vegetal sobre el planeta y la forman la mayoría de las plantas de interés agrícola e industrial.

Las micorrizas vesículo-arbusculares se conocen desde el siglo pasado, pero a pesar del interés ecológico que se deriva de su casi omnipresencia, no se les prestó demasiada atención. En los últimos 15 a 20 años, y ante la evidencia de la repercusión de esta simbiosis en nutrición vegetal, se inició un movimiento de investigación cada vez más intenso sobre el tema.

Se han descrito micorrizas VA en todos los continentes, salvo en la Antártida. Se han encontrado en Briofitas, Pteridofitas, Gimnospermas y Angiospermas. Mientras que sólo un 3 por ciento aproximadamente de las fanerógamas tienen micorrizas formadoras de manto, la gran mayoría de las especies restantes poseen micorrizas VA. Las familias de plantas en las que no se han encontrado micorrizas VA son las siguientes: a) *Pinaceae, Betulaceae* y *Fagaceae* (forman micorrizas con manto); b) *Orchidaceae* y *Ericaceae* (forman sus tipos específicos de micorrizas); c) ciertas familias que han sido descritas como no micorrizables, tales como *Chenopodiaceae, Cruciferae, Fumariaceae, Cyperaceae, Commelinaceae, Urticaceae* y *Poligonaceae.* No obstante, algunas especies de estas familias se han definido recientemente como micorrizables y no sería ilógico pensar que plantas supuestamente no micorrizables en condiciones normales puedan serlo en otras condiciones de cultivo. Existen también ciertos grupos de plantas que tienen tanto micorrizas formadoras de manto como micorrizas VA; tal es el caso de *Salicaceae, Juglandaceae, Tiliaceae, Mirtaceae, Caesalpinaceae, Juniperus, Chamaecyparis* y *Quercus.*

Aunque la ausencia de manto de micelio externo dificulta el reconocimiento de las micorrizas VA, utilizando una técnica de clarificación y tinción podemos adentrarnos en su morfología a través del examen microscópico con-

siguiente. En contraste con lo que sucede en las micorrizas formadoras de manto, la infección VA origina pocos cambios morfológicos en la raíz.

La infección se desarrolla a partir de las clamidosporas (esporas de resistencia formadas por el hongo) o bien a partir del micelio originado en una raíz previamente infectada. Las clamidosporas, que resisten condiciones adversas en el suelo, tales como el calor y la sequía, germinan cuando las circunstancias son favorables, pero los tubos de germinación producidos mueren a no ser que encuentren una raíz huésped y penetren con éxito en ella. En este caso, el tubo de germinación, o la hifa infectiva, forma un apresorio sobre la superficie de la raíz, produciéndose así la penetración del hongo, que tiene lugar normalmente entre dos células epidérmicas. (En botánica, apresorio es la protuberancia o intumescencia formada en una hifa o en el tubo germinal de una espora fúngica, destinada generalmente a adherirse al huésped durante la primera fase de la infección.) A continuación, la hifa invasora se ramifica intercelularmente, de forma rápida, en la corteza de la raíz, sin invadir endodermis, tejidos vasculares ni meristemos.

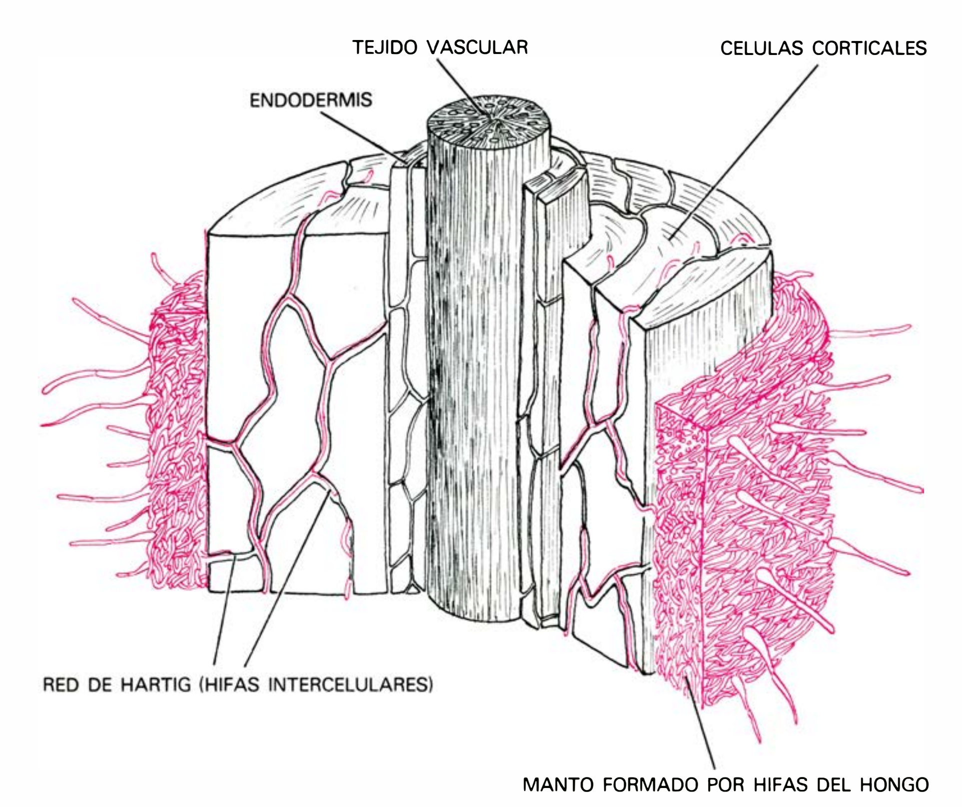
Poco tiempo después de iniciada la infección se desarrollan los arbúsculos mediante ramificación dicotómica repetida de hifas intracelulares, hasta la formación de hifas de menos de 0,2 micrometros de diámetro. Cuando se forma un arbúsculo, el almidón de la célula invadida desaparece, al tiempo que el núcleo se alarga y divide. Los arbúsculos son digeridos rápidamente y su contenido, absorbido por el huésped. Después de que los arbúsculos son digeridos, los núcleos vuelven a su tamaño normal y el almidón suele reaparecer.

Posteriormente a los arbúsculos se forman las vesículas, que son estructuras ovoides que contienen material lipídico. Estas son órganos de reserva, y, en algunos casos, su pared gruesa las asemeja a clamidosporas. Se forman intra o intercelularmente y tanto fuera como dentro de la raíz. El desarrollo de la infección en el interior de la corteza está acompañado por un crecimiento exterior de las hifas, estableciéndose posteriores puntos de entrada. Las hifas que emergen de la raíz se extienden por el suelo varios centímetros, dando lugar al micelio externo, que constituye el sistema de absorción de nutrientes. Este consta de una red tridimensional de hifas; unas, de 8-30 micrometros de diámetro que son consideradas la base permanente del micelio y, otras, más delgadas (2-7 μm), de posible función rizoidal, más efímeras que las anteriores. Sobre el micelio externo se forman grandes esporas vegetativas que van madurando hasta convertirse en clamidosporas; determinadas especies desarrollan también esporocarpos.

El parecido anatómico de las infecciones VA hizo pensar que la mayoría de ellas estaban causadas por el mismo hongo; de hecho, hasta la revisión de Mosse en 1973 se aceptaba que los endofitos VA pertenecían en su mayoría al género *Endogone*, del cual se habían descrito diversidad de tipos o razas. De acuerdo con las clasificaciones clásicas, todas las especies de *Endogone* formaban esporocarpos, pero ante el descubrimiento de esporas de resistencia ectocárpicas unidas al micelio externo de raíces micorrizadas se planteó la necesidad de revisar la taxonomía de estos hongos. En este sentido, se llevaron a cabo una serie de investigaciones que han permitido encuadrar actualmente a los hongos VA en la familia *Endogonaceae* de los *Mucorales* (ficomicetos), agrupados en cuatro géneros: *Glomus*, *Sclerocystis*, *Gigaspora* y *Acaulospora*, ninguno de los cuales ha podido ser aislado en cultivo puro.

Los hongos que forman micorrizas VA tienen un espectro de huéspedes extremadamente amplio, lo que obliga a catalogarlos como inespecíficos. Pero sí se registran diferencias en el grado de susceptibilidad del huésped y en la adaptabilidad del hongo a determinadas condiciones. La existencia, por ejemplo, de hongos más adaptados a especies forestales y otros a cultivos agrícolas es un hecho comprobado; así mismo, se sabe que el *p*H de un suelo es un factor determinante de la presencia y efectividad de ciertos tipos de esporas. Esto parece indicar que sí hay marcadas diferencias en la facilidad e intensidad con que los endofitos infectan, se desarrollan y operan en distintos huéspedes y bajo diferentes condiciones ecológicas. Estas consideraciones permiten concluir que existe cierta especificidad en micorrizas VA, en términos de efectividad de la simbiosis, aunque ésta parece depender más de la interacción con un tipo de suelo y condiciones de cultivo que no con un huésped particular.

El potencial infectivo de un suelo viene determinado por la cantidad de esporas presentes y por la intensidad de la infección de las raíces que contiene. Se sabe que tales parámetros están influenciados por las condiciones ambientales. La mayoría de las esporas y



ECTOMICORRIZAS o micorrizas formadoras de manto. Se caracterizan porque el hongo que las origina se desarrolla en la superficie de la raíz formando un auténtico manto de hifas que la cubren. En el interior de la raíz (corteza) el hongo se desarrolla intercelularmente constituyendo la red de Hartig, tal como se observa en este esquema de un corte transversal de una ectomicorriza. Aproximadamente un 3 por ciento de las plantas superiores forman este tipo de micorrizas, siendo la mayoría especies de interés forestal, entre las cuales merece la pena destacar: pino, abeto, abedul, haya, roble, eucaliptus, etcétera.

de las raíces micorrizadas se encuentran en la capa superficial del suelo, lo que quizá sea simplemente reflejo de una mayor densidad de raíces en la capa arable de los suelos, aunque no se debe olvidar que la escasez de oxígeno molecular en capas más profundas puede ser el factor condicionante.

Las micorrizas VA estimulan el crecimiento, desarrollo y nutrición de las plantas, especialmente en suelos de baja y moderada fertilidad. Los estudios llevados a cabo han puesto de manifiesto que dichos efectos se deben a que la micorriza mejora sustancialmente la absorción de nutrientes y agua por la planta y que el principal nutriente implicado es el fósforo.

Se sabe que la mayor parte de los suelos naturales tienen un bajo contenido en fosfato asimilable, e incluso la mayoría de los suelos arables productivos necesitan un aporte considerable de fertilizante fosforado para mantener su fertilidad. En efecto, el 95-99 por ciento del fósforo de un suelo está integrado en compuestos orgánicos o inorgánicos insolubles. De otro lado, se conoce que el ritmo de absorción de los iones fosfato por la planta es superior al de desplazamiento de dichos iones desde el suelo no rizosférico hacia la raíz. Ello condiciona que se forme una zona de agotamiento del elemento en la rizosfera. Esta zona de agotamiento, que ha podido ser puesta de manifiesto por autorradiografía, es la base que justifica que el PO_4^{-3} sea factor limitante del crecimiento de las plantas en gran número de suelos. Efectivamente, el desplazamiento del ion fosfato hacia la raíz tiene lugar por difusión; en su camino hacia la planta se fija fácilmente a arcillas y coloides del suelo por medio de combinaciones insolubles con calcio, hierro o aluminio.

Los mecanismos propuestos para explicar la mayor capacidad de absorción de fósforo por plantas micorrizadas están basados en las siguientes causas: *a*) que la micorrización induzca cambios morfológicos en la planta; *b*) que la micorrización induzca cambios fisiológicos, lo que provocaría un incremento de la capacidad de la superficie de la raíz de la planta para absorber fósforo; *c*) que la micorrización proporcione una superficie de absorción adicional (hifas del hongo), o más eficaz; *d*) que las hifas o las raíces micorrizadas tengan capacidad para solubilizar fuentes de fósforo no disponibles para raíces no infectadas y *e*) que la raíz micorrizada tenga más longevidad que la que no lo está.

La posibilidad *d*) ha sido objeto de especial investigación y controversia. De un lado, muchos experimentos indican que las plantas micorrizadas crecen mejor que las no micorrizadas en suelos enriquecidos con formas difícilmente solubles de fósforo, tales como apatito, fitato y fosfatos de Ca, Al o Fe. De otro lado, en micorrizas formadoras de manto, se ha descrito la presencia de fosfatasas de superficie. Todo esto hizo pensar en una posible solubilización de fosfatos por las hifas de las micorrizas VA. Sin embargo, ensayos con fósforo radiactivo (P^{32}) pusieron de manifiesto que, tanto las plantas micorrizadas como las no micorrizadas, tomaban el fósforo de la misma fuente: el depósito ("pool") lábil de fosfato soluble. La absorción más eficiente por las raíces micorrizadas causa que se estimule la disociación química del fosfato insoluble para reponer el soluble, que están captando las hifas de la micorriza, y mantener así el equilibrio fosfato insoluble-fosfato soluble. Ello puede justificar la respuesta de las plantas micorrizadas a la adición de fosfato de roca. No obstante, es aconsejable continuar la investigación sobre la posible utilización de formas no asimilables de fosfato por las micorrizas VA, en ciertos suelos o condiciones todavía por estudiar.

La eficacia de las raíces en la absorción de fosfatos se ha valorado midiendo la cantidad de P^{32} captado por unidad de longitud de raíz. Se ha estimado que este parámetro es cuatro veces mayor para las raíces micorrizadas que para las no micorrizadas, lo que quiere decir que las raíces micorrizadas absorben los fosfatos mucho más eficazmente que las que no lo están. Las explicaciones basadas solamente en cambios morfológicos de la raíz (hipótesis *a*) no justifican ese considerable incremento en la eficacia, máxime cuando se conoce que los cambios morfológicos que aparentemente originan las micorrizas VA son insignificantes. Por otra parte, las razones basadas en suponer una mayor longevidad de

MICORRIZAS VESICULO-ARBUSCULARES: constituyen el tipo de micorrizas más ampliamente distribuido en la naturaleza, ya que las forman, aproximadamente, el 96 por ciento de las plantas existentes sobre la Tierra, entre las que se encuentran especies de gran interés agrícola e industrial. No forman manto externo de hifas y se desarrollan en el interior de la raíz, inter e intracelularmente, dando lugar a los elementos morfológicos típicos de esta infección: los arbúsculos y las vesículas. En el esquema se muestra un corte longitudinal de una micorriza VA, en el cual se aprecian los mencionados elementos.

TIPOS DE MICORRIZAS

DENOMINACION CLASICA	DENOMINACION ACTUAL	CARACTERISTICAS	PLANTA HUESPED	HONGOS QUE LA FORMAN
Ectotróficas	Formadoras de "manto" (sheating)	– Forman "manto" que cubre la raíz. – Hifas sólo intercelulares que forman la red de Hartig. – Hongo de micelio septado.	*Betulaceae* *Fagaceae* *Pinaceae* *Eucaliptus*	*Agaricaceae* *Boletaceae* y otros
Endotróficas	Vesículo-arbusculares (VA)	– Desarrollo mayoritario del hongo dentro de la raíz. – Hifas externas no formadoras de manto. – Micelio no septado, salvo en hifas viejas. – Hifas inter e intracelulares: las intercelulares no forman red de Hartig, las intracelulares forman arbúsculos y vesículas.	Se han encontrado en la mayoría de las plantas que viven sobre la corteza terrestre.	Ficomicetos microscópicos pertenecientes a la Familia *Endogonaceae.*
	Ericáceas: Ericoides	– Rudimento de manto. – Hifas inter e intracelulares: las intracelulares forman masas compactas que pueden ser lisadas o digeridas. – No se forman vesículas ni arbúsculos.	*Ericaceae* *Epacridaceae* *Empetraceae*	Ascomicetos
(Ectendotróficas)	Ericáceas: Arbutoides	– Forman manto. – Hifas intra e intercelulares: las intercelulares no forman red de Hartig.	*Ericaceae* { *Arbutus*, *Arctostaphylos* } *Pyrolaceae* *Monotropaceae*	*Boletus*
	Orquidáceas	– La planta huésped tiene un período de su ciclo de vida heterótrofo durante el cual, para sobrevivir, necesita ser infectada por un hongo micorrízico. – La infección del huésped por el hongo puede evolucionar a micorriza o parasitismo.	*Orchidaceae*	Basidiomicetos

EXISTEN DIVERSOS TIPOS de micorrizas que difieren en sus características morfológicas y en su significado ecológico. En las micorrizas formadoras de manto, vesículo-arbusculares y ericáceas, el hongo ayuda a la planta a obtener sus nutrientes minerales del suelo y, a cambio, la planta cede al hongo hidratos de carbono. La relación planta-hongo en estos casos se considera una simbiosis mutualística. Las micorrizas de las orquídeas presentan unas características fisiológicas diferentes, por lo que han sido objeto de especial interés. Las orquídeas son heterótrofas en algún período de su ciclo de vida y, durante éste, se muestran totalmente dependientes de las micorrizas, obteniendo hidratos de carbono de otras plantas vecinas por medio del hongo.

las raíces micorrizadas no parecen aceptables, si se tiene en cuenta que las raíces no micorrizadas son capaces de absorber fósforo durante largos períodos de tiempo.

La hipótesis *c*), a saber, que la micorrización proporcione una superficie de absorción adicional, se considera hoy día la más acertada. En efecto, se acepta que el papel fundamental de las micorrizas estriba en que las hifas externas del hongo extienden el campo de absorción de la planta más lejos de la zona de agotamiento que rodea la raíz, de tal manera que la red de hifas externas permite a la raíz incrementar su superficie de absorción y explorar un volumen de suelo superior al que pueden utilizar las plantas no micorrizadas. Esto parece estar confirmado por ciertos trabajos en los cuales se detectó captación de P^{32} por raíces micorrizadas, cuando se colocó el isótopo a una distancia no accesible a raíces no micorrizadas. Concretamente, se ha encontrado que las hifas de *Glomus* pueden extender la zona de captación de fosfato hasta, al menos, siete centímetros de la superficie de la raíz. No hay que olvidar, además, que un centímetro de raíz micorrizada puede tener hasta 80 centímetros de hifas externas.

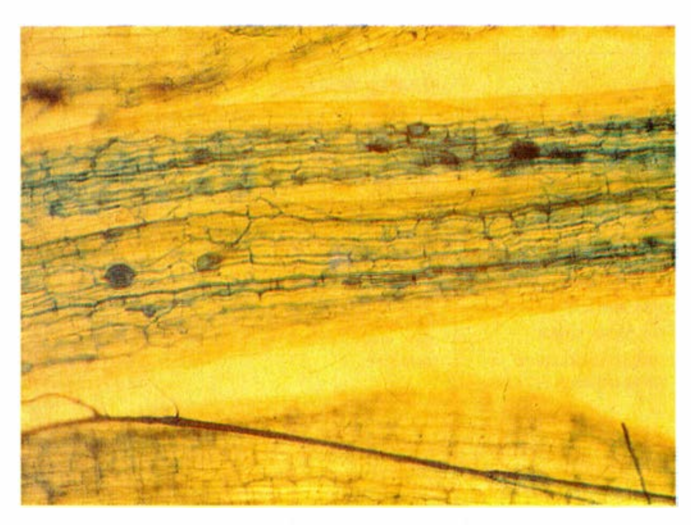
Utilizando P^{32} se ha puesto de manifiesto que el incremento en la captación de fosfato por las hifas depende más de su posición, longitud y número, que de alguna propiedad especial de su superficie que agilice la captación. Estas apreciaciones parecen descartar la hipótesis *b*): la micorrización no induciría pues cambios fisiológicos que justifiquen la mayor absorción de fósforo.

De otro lado, se ha calculado que el flujo de fosfato en el interior de las hifas es del orden de 10^{-8} a 10^{-9} moles por cm^{-2} por seg^{-1}, habiéndose propuesto varios mecanismos para tratar de explicarlo. Diversas observaciones indican la existencia de ciclosis o corrientes protoplasmáticas bidireccionales en el interior de la hifa, aceptándose que éstas desempeñan un papel importante en la traslocación del fosfato.

Este modelo de traslocación ha podido ser completado gracias a trabajos histológicos, en los cuales se ha puesto de manifiesto la existencia de gránulos de polifosfato en las vacuolas de hifas y arbúsculos, particularmente cuando la infección está en sus estadios jóvenes y vigorosos. Basándose en estas observaciones, se ha propuesto que el ion fosfato se trasloca como gránulos de polifosfato, que se van depositando en las vacuolas, incrementando su tamaño o disminuyéndolo a medida que se van utilizando. La operatividad de este mecanismo está asegurada con la existencia de polifosfatasas ácidas, y sobre todo alcalinas, localizadas en las vacuolas de arbúsculos maduros e hifas intercelulares, descrita por varios autores.

En cuanto a la transferencia de fosfatos del hongo al huésped, se acepta, desde hace tiempo, que tiene lugar en los arbúsculos, los cuales degeneran y son digeridos, liberándose el fósforo que contienen. Observaciones a nivel estructural confirman este hecho; se ha visto así que los arbúsculos se van formando y degenerando, calculándoseles una vida media de 7 a 11 días. No hay que descartar, sin embargo, la posibilidad de que ocurra una transferencia de fosfato, y otros materiales, a través de los plasmalemas fúngico y del huésped. En efecto, la transferencia puede tener lugar en otras partes del micelio interno y no sólo en los arbúsculos (intracelulares). Existen varias razones para pensar así; de hecho, en micorrizas formadoras de manto, hay intercambio de fosfato sin que ocurra la penetración intracelular del micelio. Asimismo, se han detectado respuestas a la infección VA antes de la formación de arbúsculos. Se puede concluir, pues, que el sitio fundamental de transferencia es el arbúsculo, pero puede que no lo sea exclusivamente.

Por otro lado, está demostrado que





HONGOS CAUSANTES DE LAS MICORRIZAS V-A. Originan escasos cambios morfológicos en la raíz; para detectar su presencia, se procede a la clarificación de los tejidos radicales y a la posterior tinción de las estructuras fúngicas, lo que permite observar los elementos típicos de la simbiosis VA. Las fotografías muestran un aspecto general de la infección VA en una raíz de *Medicago sativa* (alfalfa), así como detalles más concretos de vesículas.

los carbohidratos procedentes del fotosintato de la planta son transferidos al hongo, y que llegan hasta su micelio externo. Aunque no se conoce el mecanismo de la transferencia, se acepta que ésta tiene también lugar en los arbúsculos.

Finalmente, cabe considerar que se han llevado a cabo muchos trabajos para determinar si las micorrizas VA favorecen la captación de otros nutrientes. En algunos de tales trabajos, parece indicarse la captación de Zn, S, K y Ca por micorrizas VA. Sin embargo, existen contradicciones para estos y otros nutrientes, por lo que no se pueden generalizar estas apreciaciones. Probablemente, lo que ocurra sea que, por constituir el fósforo un factor limitante del crecimiento, las plantas micorrizadas están más equilibradas fisiológicamente, lo que puede condicionar una mejor absorción de otros nutrientes. De igual modo, se sugirió que las micorrizas VA estimulaban la captación de agua, lo que resultaría de gran importancia ecológica en suelos áridos, pero luego se puso en cuestión el fenómeno y necesita más estudio antes de elaborar conclusiones definitivas.

Diversos factores ecológicos pueden afectar el desarrollo y actividad de las micorrizas. La luz es un factor fundamental en la infección VA. En plantas colocadas en ligera penumbra, la infección no sufre alteraciones de relieve. Por el contrario, cuando se las somete a grandes sombras, la infección se reduce drásticamente y la producción de esporas baja en un 80 por ciento. Igualmente, las bajas temperaturas reducen también la infección y la producción de esporas.

El porcentaje de infección y el número de esporas se reducen, por lo general, cando se aplican fertilizantes fosforados y nitrogenados. Podemos afirmar que las micorrizas VA son más persistentes en suelos de baja y moderada fertilidad, aunque existan abundantes excepciones a tal generalización. Interesa, en este sentido, profundizar algo más en el efecto del PO_4^{-3}, en particular, sobre la micorrización. Adiciones crecientes de este nutriente reducen la infección, así como la formación de esporocarpos. Se ha podido demostrar, mediante ensayos en los que se aplica foliarmente fosfato soluble, que la concentración del ion dentro de la planta tiene más influencia en la reducción de la infección que el existente en el suelo.

Ciertas observaciones han sugerido que las diferencias en el grado de infección asociadas con la fertilidad del suelo están relacionadas con el ritmo de crecimiento de la raíz. Algunos autores opinan que las raíces vigorosas en crecimiento activo son difícilmente infectadas, y concluyen que cualquier factor que cause un crecimiento lento de la raíz, o que reduzca la proporción de tejido radical en crecimiento activo, tiende a incrementar la infección. Puesto que se sabe que las auxinas y el etileno, fundamentalmente, y la acción combinada de éstas con citoquininas y giberelinas controlan la formación y desarrollo de las raíces, es lógico pensar que estas sustancias hormonales intervienen de una manera destacada en la



LA FORMA DE RESISTENCIA de los hongos VA es la espora (clamidospora). La forma, contenido, color y modo de unión a la hifa de sustentación son los criterios usados en la diferenciación de los distintos géneros y especies de la familia *Endogonaceae*, a la que pertenecen los hongos microscópicos responsables de estas micorrizas. Algunas especies forman esporocarpos, en cuyo interior se encuentran las esporas. En la fotografía superior se muestra la espora de tipo "yelow vacuolate" (de vacuolas amarillas) que corresponde al género *Glomus mosseae*, integrada en un esporocarpo perteneciente a la misma especie.

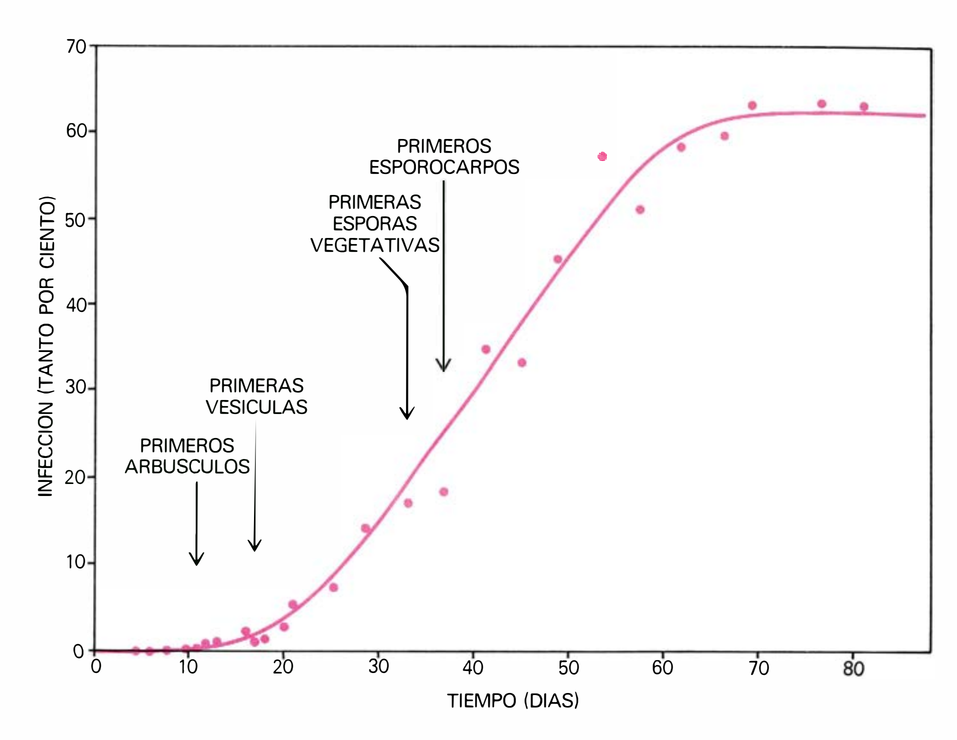
formación de micorrizas VA, hecho que ha podido ser demostrado en nuestro laboratorio.

En las micorrizas formadoras de manto se ha demostrado que los hongos que las originan producen auxinas, y que éstas son rápidamente absorbidas por las raíces de las plantas huéspedes, originándose cambios morfológicos que facilitan el establecimiento de la simbiosis. De otro lado, también se ha puesto de manifiesto que estos hongos producen citoquininas y giberelinas, sin que se sepa exactamente la función de éstas en el proceso de infección.

La posibilidad de llevar a cabo estas investigaciones en hongos VA ha entrañado problemas que arrancan de la dificultad de desarrollar estos hongos en cultivo puro. No obstante, utilizando técnicas puestas a punto por nosotros que permiten un crecimiento inicial del hongo, hemos comprobado la producción de auxinas, giberelinas y citoquininas por éstos.

Varios trabajos experimentales han apuntado la existencia de interacciones entre hongos de las micorrizas y otros microorganismos de la rizosfera. Hay que tener en cuenta que los exudados radicales de las plantas micorrizadas difieren de los de las no micorrizadas, no sólo porque cambia el estado nutricional de la planta huésped, sino también debido a la gran cantidad de tejido del hongo presente en la raíz micorrizada, que puede afectar más directamente a los exudados. En el caso de las micorrizas formadoras de manto, se sabe que éstas inducen cambios cualitativos y cuantitativos en la microflora del suelo; a su vez, se ha descrito que *Azotobacter* y otros microorganismos afectan de una manera positiva la formación de esta simbiosis. Tal efecto se debe, probablemente, a las sustancias extracelulares que los microorganismos liberan, entre las cuales pueden desempeñar un importante papel las fitohormonas, cuya producción es un fenómeno común de los microorganismos de la rizosfera. Igualmente, se ha señalado un efecto beneficioso de *Pseudomonas* sp. en el establecimiento de las micorrizas VA, así como la cooperación entre los hongos VA y bacterias solubilizadoras de fosfatos en cuanto a sus efectos sobre el crecimiento y nutrición de las plantas y también en el establecimiento de la simbiosis. De igual forma se han sugerido interacciones entre *Azotobacter* y hongos VA.

Desde el punto de vista de su repercusión en agricultura, las interacciones entre hongos VA y *Rhizobium* sp. evidencian un enorme interés, en concreto por sus efectos sobre el crecimiento y nutrición de leguminosas. Nuestros estudios han puesto de manifiesto la existencia de interacciones de dichos microorganismos, no sólo a nivel nutritivo, sino también a nivel de los mecanismos que intervienen en la formación de estas simbiosis con las plantas. Finalmente, cabe señalar, aunque sólo sea de paso, la influencia negativa que sobre la micorrización parecen ejercer ciertos compuestos normalmente usados en agricultura como son los pesticidas.



DINAMICA DEL DESARROLLO DE LA INFECCION VA en una planta hospedadora adecuada. Aunque con peculiaridades y matices propios de las especies implicadas y factores ecológicos controlantes, se caracteriza por seguir un modelo en 3 fases: fase *lag*, en la que tiene lugar la germinación de las esporas y comienza la colonización de las raíces por las hifas; a continuación, sigue una fase de desarrollo intensivo de la infección, y, por último, se llega a la fase de constancia en la cual no varía la proporción entre raíces micorrizadas y no micorrizadas. En la gráfica se muestra la dinámica del proceso de infección en raíces de *Medicago sativa* por el hongo de la endomicorriza *Glomus mosseae*.

Se entiende por micotrofismo la alimentación por medio de hongos. En este sentido, se dice que ciertas especies son más "micotróficas" que otras cuando obtienen más beneficio de las micorrizas. La dependencia de las micorrizas se define como "el grado en que una planta depende de la condición de estar micorrizada para producir su máximo crecimiento y rendimiento a un nivel dado de fertilidad del suelo". En un extremo de esta graduación están las plantas no micorrizables, absolutamente no-dependientes de las micorrizas, y en el otro se sitúan ciertas variedades de *Citrus* que son "altamente micorriza-dependientes", ya que ni siquiera una fertilidad elevada les proporciona los beneficios de la micorrización. La mayoría de las plantas se encuentran entre estos extremos, y para cada especie y situación se supone la existencia de un nivel de fertilidad (fundamentalmente de fosfato asimilable) al cual crecen igual las plantas micorrizadas que las que no lo están.

Generalmente, las plantas con alta demanda de fósforo (como las leguminosas) o pobre sistema radical (cebolla, patata) responden mejor a la micorrización. Se ha indicado que la capacidad de las plantas para crecer en suelos que tienen muy poco fósforo disponible puede depender del desarrollo de los pelos radicales; así, plantas con pocos o cortos pelos radicales dependerán más de la formación de micorrizas que las dotadas con pelos bien desarrollados. Al parecer, sin embargo, se hallan implicados otros factores, además de la cantidad y tamaño de los pelos, pues plantas como *Paspalum notatum* y *Centrosema pubescens,* con largos pelos radicales, no toman eficientemente el fósforo del suelo, a no ser que estén micorrizadas.

Como se indicó anteriormente, la colonización de la superficie terrestre por las plantas tuvo lugar hace unos 400 millones de años, entre el Silúrico y el Devónico. La planta fósil más antigua que ha podido ser estudiada pertenece al género *Rhynia* y data de hace unos 370 millones de años. Se han obtenido microfotografías de las "raíces" de esta planta, en las cuales se aprecia claramente la presencia de una

formación fúngica en su interior de gran similitud con las actuales especies de *Glomus*. Algunos autores consideran que, en ciertos casos, tales asociaciones hongo-"raíz" eran ya micorrizas.

Es muy probable que las primeras plantas que evolucionaron y se establecieron en un ecosistema terrestre encontraran un suelo ya colonizado por poblaciones microbianas, entre las cuales posiblemente hubiera cianobacterias y bacterias fijadoras de N_2. Por esta razón, es probable que no fuera el nitrógeno el factor limitante, sino el fósforo, tal como ocurre actualmente en la colonización de nuevos hábitats. En tales condiciones, cualquier asociación planta-microorganismo que mejorara la extracción de fósforo del medio se seleccionaría y perpetuaría. Se ha sugerido que ahí podría radicar el origen de las micorrizas VA. El micotrofismo, por tanto, no es un proceso nuevo, sino que nació con la aparición de plantas con raíces.

Se acepta comúnmente que las plantas vasculares se originaron a partir de las algas verdes. En este sentido, se ha propuesto una hipótesis según la cual, para que el alga semiacuática pudiera superar los problemas de desecación y nutrición que suponía el paso a un hábitat terrestre, debieron asociarse simbióticamente con un hongo acuático. De acuerdo con esta hipótesis, las plantas serían el producto de tal simbiosis, la cual permitiría la colonización de la tierra durante el Silúrico y el Devónico. La conclusión de esta hipótesis es taxativa: las plantas terrestres nunca tuvieron independencia, porque si la hubieran tenido jamás hubieran colonizado la tierra. La simbiosis fue, posiblemente, de tipo endotrófico, modelo de asociación mantenida hoy día por casi todas las plantas en forma de micorrizas VA. Diremos, pues, que las micorrizas se originaron prácticamente con las plantas, permitieron su evolución y a su vez evolucionaron con ellas y con ellas se han perpetuado.

Hoy se reconoce que los microorganismos que viven en el suelo desarrollan en su hábitat natural actividades de gran interés por su repercusión directa en el crecimiento y nutrición de las plantas, e, indirectamente, en la nutrición de los animales.

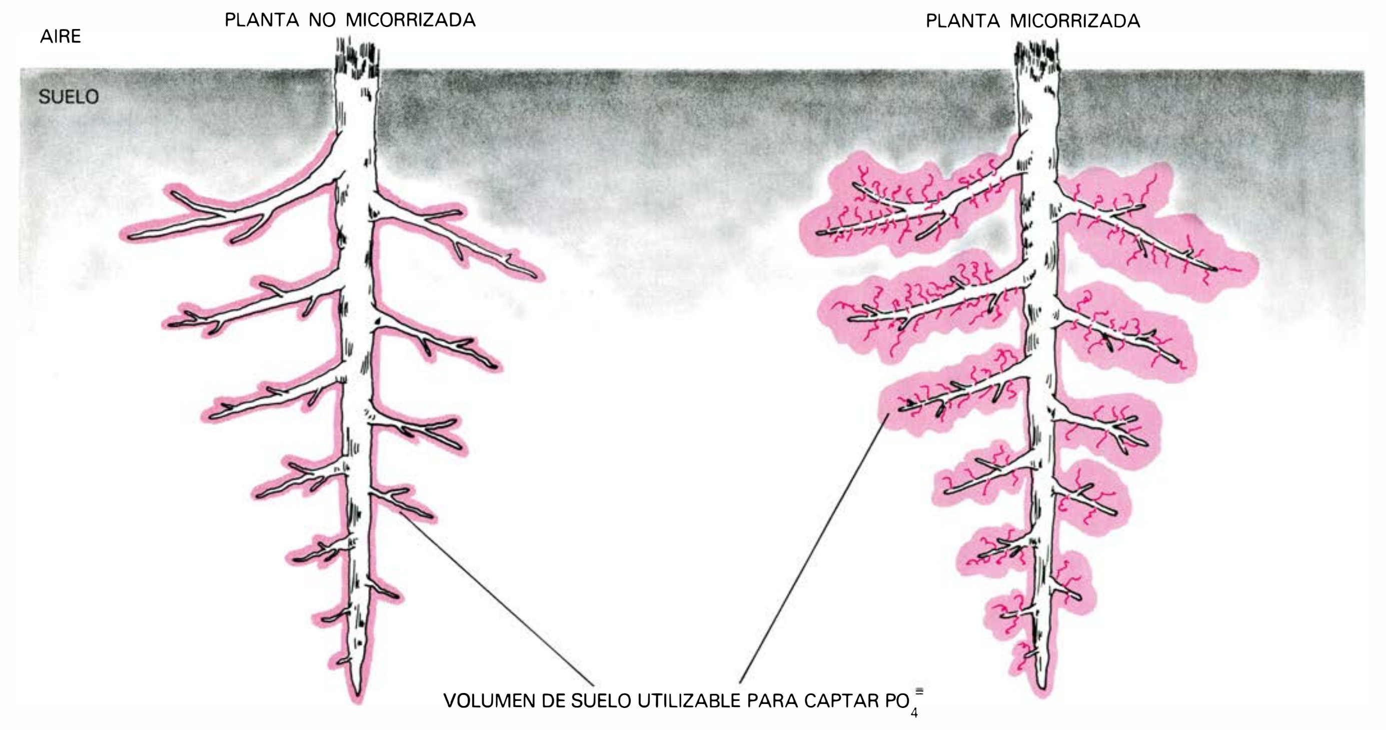
Es un hecho incuestionable que la necesidad de incrementar la calidad y cantidad de las cosechas agrícolas, ante la creciente demanda de alimentos, condiciona el progresivo agotamiento de los suelos, lo que, a su vez, induce a aumentar el consumo de fertilizantes químicos, de elevado coste. Estos hechos han animado a investigar la posibilidad de manipular las poblaciones de microorganismos que habitan en el suelo y que intervienen en los ciclos de los nutrientes, con el fin de mejorar el contenido en formas asimilables por las plantas cultivadas. La idea, lógicamente, es incrementar la producción vegetal y reducir el consumo de fertilizantes de elevado precio y contaminantes, a la larga, de los ecosistemas.

Actualmente, la atención se ha centrado en el estudio de microorganismos capaces de formar asociaciones simbióticas mutualísticas con las plantas. Así, éstas se benefician directamente de la acción de los microorganismos, que, a su vez, encuentran en la planta un nicho ecológico protegido y fuentes de carbono asimilable. Tal es el caso de la simbiosis *Rhizobium*-leguminosa que, como se sabe, representa un importante aporte de nitrógeno a esas plantas y, a través de ellas, a la biosfera.

Igualmente, se hallan en estudio las posibilidades de utilizar las micorrizas como fertilizantes biológicos. En este sentido, se han llevado a cabo ensayos, generalmente en invernadero, en los que se ha puesto de manifiesto la ventaja de la inoculación con hongos formadores de micorrizas VA. Aunque estos hongos se encuentran de modo natural en los suelos, puede ocurrir, debido a diversas circunstancias tales como empleo abusivo de fertilizantes químicos, productos fitosanitarios, etcétera, que falten, escaseen, o incluso que, existiendo en número adecuado, sean poco efectivos. Es en estos casos en los que se espera que la introducción de endofitos VA muestre especial eficacia, por lo que en la mayoría de los ensayos llevados a cabo en suelos naturales se han usado esterilizantes para eliminar los endofitos nativos. No obstante, se han descrito algunos ensayos en campo sobre suelo no tratado, en los cuales se obtuvieron incrementos de hasta un 150 por ciento en la cosecha como respuesta a los tratamientos con endofitos introducidos frente a los que sólo poseían los nativos del suelo en cuestión.

Sin embargo, a diferencia de lo que ocurre en el caso de la simbiosis *Rhizobium*-leguminosa, en la cual el N_2 fijado procede de la atmósfera, las micorrizas mejoran la captación por la planta del fósforo ya existente en el suelo, por lo cual, a no ser que se

EL FOSFORO ES UN NUTRIENTE FUNDAMENTAL para el crecimiento vegetal, ya que forma parte de moléculas de gran importancia en la economía celular, como son los ácidos nucleicos, así como de los compuestos encargados de la captación, almacenamiento y transferencia de la energía, el ATP, por ejemplo. Este elemento nutritivo es factor limitante del crecimiento de las plantas en la mayoría de los suelos, al encontrarse formando parte de fosfatos orgánicos o minerales insolubles no utilizables por el vegetal. En el ciclo del fósforo en la naturaleza, según indica el esquema, se aprecia que los microorganismos del suelo intervienen en la solubilización y mineralización de fosfatos, aunque hoy se admite que la acción más decisiva es la que realiza un grupo muy especial de microorganismos: los hongos que forman las micorrizas, cuyas hifas externas canalizan el fósforo desde el suelo hasta la planta.



HIFAS EXTERNAS del hongo de la micorriza. Debido a su longitud y distribución, permiten a la planta explorar un volumen de suelo superior al que pueden utilizar las plantas no micorrizadas, aumentando su campo de absorción de fósforo, lo que reviste gran interés en la nutrición vegetal.

adicione dicho elemento, se producirá un empobrecimiento en fosfato del medio. Esto quiere decir que las micorrizas no pueden sustituir de forma total a los fertilizantes químicos, sino que ambos tipos de fertilizantes deben usarse conjuntamente, de suerte que la adición de fósforo asegure el mantenimiento del nivel de dicho elemento en el suelo y no inhiba el proceso de micorrización, sino que, por el contrario, coopere con la micorriza para conseguir una producción vegetal óptima.

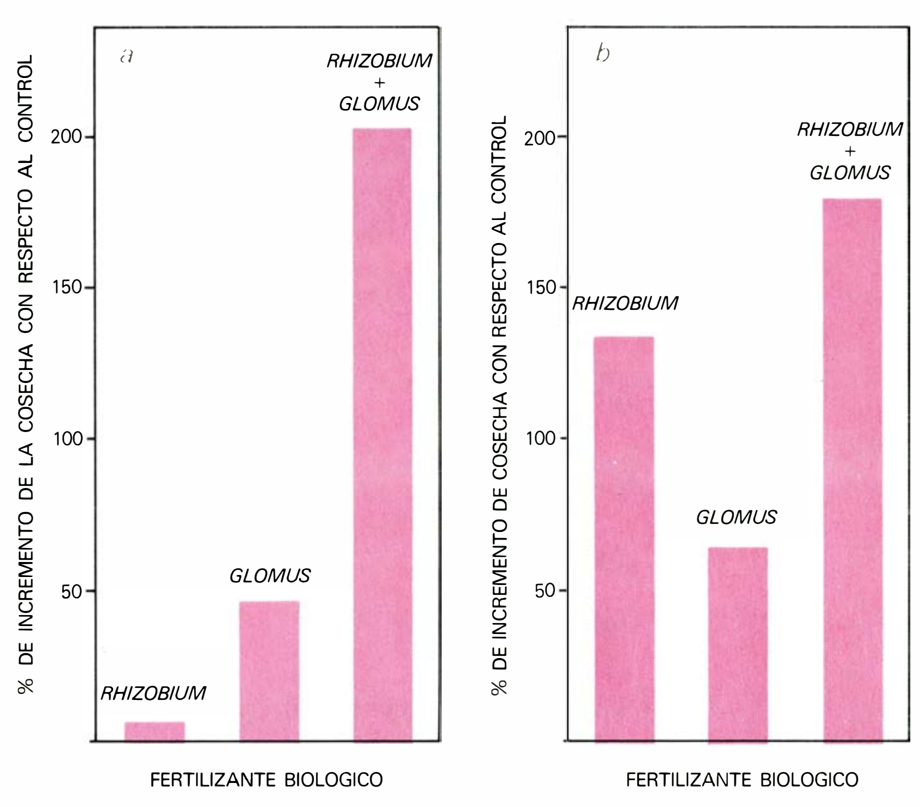
Ello puede alcanzarse de dos maneras distintas. Consiste la primera en la aplicación de fosfatos que, aunque difícilmente solubles, son más baratos, como por ejemplo el fosfato de roca. Se sabe que éste no inhibe la formación ni los efectos de las micorrizas y que incluso puede mejorar la actividad de éstas, ya que aunque las plantas micorrizadas no pueden captarlo directamente, la absorción más eficiente del fósforo soluble por parte de las hifas del hongo estimula la disociación química del fosfato insoluble para reponer el soluble que se está consumiendo y mantener así el equilibrio. La otra posibilidad consiste en aplicar formas solubles de fosfato, si bien en una cantidad y momento tales que no inhiban la constitución de la simbiosis. Aunque las dosis elevadas afectan negativamente el desarrollo de las micorrizas, dosis más bajas pueden no ser perjudiciales, y hasta cooperar con ellas, consiguiéndose así mejores resultados.

Cualquiera de estas dos posibilidades es válida para mantener un nivel adecuado de fosfato en el suelo y permitir que las plantas se beneficien de los efectos de la micorrización. La elección de una de estas formas va a depender de las características del suelo (capacidad fijadora de fosfato, *p*H, etcétera) y de la planta a utilizar. Por ello, cuando se quiera llevar a cabo un ensayo de fertilización biológica en campo, debe desarrollarse antes una investigación que permita establecer la necesidad de inocular, seleccionar los microorganismos adecuados y estudiar su interacción con los fertilizantes químicos. Siguiendo estas pautas, nosotros hemos llevado a cabo una serie de investigaciones sobre el tema. Hemos utilizado leguminosas por ser estas plantas capaces de formar los dos tipos mencionados de simbiosis con microorganismos.

Se pretendía estudiar las posibilidades de practicar con éxito unos ensayos de fertilización biológica en los suelos relativamente fértiles de la Vega de Granada. Se trata de suelos neutroalcalinos y fuertemente fijadores de fósforo. Esta circunstancia, unida al hecho de que reciben considerables cantidades de fertilizantes fosforados, hace que el contenido en fósforo total sea muy elevado, y, bajo, el fósforo asimilable. Los ensayos previos llevados a cabo en invernadero pusieron de manifiesto que la introducción de hongos de la micorriza VA mejoraba el crecimiento, nutrición y nodulación de las plantas.

Los endofitos VA introducidos cooperaron de modo activo con los indígenas. En concreto, se estimó que la micorrización provocaba un 160 por ciento, aproximadamente, de incremento de cosecha, siendo los hongos VA introducidos responsables de un 100 por ciento y, los nativos, del 60 por ciento restante. Desde el punto de vista práctico, cabe destacar que la micorrización tuvo un efecto similar al de la dosis de fosfato soluble correspondiente a la empleada por los agricultores de la región.

De todo lo que antecede se dedujo la posibilidad de llevar a cabo con éxito ensayos de inoculación en campo con *Rhizobium* y hongos de la micorriza. Los ensayos se desarrollaron en parcelas distintas del suelo en cuestión, que diferían en algunas propiedades químicas, como el contenido en fósforo asimilable (17,6 partes por millón [ppm] en la parcela *b* frente a 9,2 ppm en la *a*). Al mismo tiempo, se procedió al recuento de esporas de los hongos VA; se evidenció que ambas parcelas poseían un número bajo de esporas, pero que la parcela *b* contenía tres veces más ($75 \pm 9{,}3$) que la *a* ($24 \pm 0{,}7$). La especie de hongo VA predominante fue *Glomus mosseae,* por lo que se utilizó este microorganismo como inoculante. Por otro lado, es obvio que por ser la alfalfa (*Medicago sativa*) la planta a utilizar en los ensayos, tenía que emplearse *Rhizobium meliloti* co-



***RHIZOBIUM* Y EL HONGO DE LA MICORRIZA VA *(Glomus mosseae)* capaces, respectivamente, de fijar (N_2) atmosférico y de captar fósforo para la planta, se aplicaron como "fertilizantes biológicos" a *Medicago sativa* (alfalfa) crecida en condiciones naturales. Las plantas respondieron a la inoculación del hongo de la micorriza y, más aún, a la aplicación del doble inóculo, pero sólo hubo respuesta positiva a la introducción del microorganismo *Rhizobium* cuando el suelo poseía cierto nivel de fosfato asimilable, o un nivel de micorrización adecuado, caso del ensayo registrado en la ilustración de la derecha. Estos experimentos de fertilización biológica en campo fueron desarrollados por los autores de este artículo.**

mo especie capaz de nodular y fijar nitrógeno en simbiosis con esta planta.

De los resultados de los ensayos de campo se dedujeron las siguientes conclusiones: *1*) La inoculación con *Glomus* fue eficaz en todas las situaciones estudiadas. *2*) La inoculación con solo *Rhizobium* no fue eficaz en la parcela *a* debido, probablemente, a que el fósforo actuó de factor limitante del crecimiento y la nodulación de las plantas, ya que sí hubo respuesta cuando se inocularon las plantas con el hongo de la micorriza. En la parcela *b*, sin embargo, hubo respuesta ante la aplicación de *Rhizobium*, ya que el nivel de fósforo soluble en el suelo era superior y, además, el número superior de esporas de hongos VA nativos permitiría un mayor grado de micorrización de las plantas con respecto a las de la parcela *a*. *3*) En todos los casos, la inoculación de *Rhizobium* + *Glomus* mostró gran eficacia incrementando la cosecha en, aproximadamente, un 200 por ciento sobre los testigos.

Como se apuntó anteriormente, es más probable que la introducción de endofitos VA tenga éxito en suelos de baja fertilidad, pero nuestros estudios indican que la micorrización también puede ser aconsejable en suelos relativamente fértiles, como los usados en estos ensayos. Conviene añadir que en los experimentos de campo no se adicionaron fertilizantes químicos y que las labores fueron practicadas por los agricultores en su forma tradicional.

Es evidente que los resultados obtenidos muestran unas perspectivas alentadoras. Sin embargo, aún se necesita mucha más investigación en este sentido para generalizar los resultados. Por un lado, hay que estudiar, en cada caso concreto, diversos aspectos de la ecología de las micorrizas que permitan establecer más bases científicas de la inoculación. Por otro, mientras que la obtención de inoculantes de *Rhizobium* es un proceso rutinario e industrializable, el conseguir inóculo masivo de micorriza muestra actualmente muchas dificultades, ya que el hongo, simbionte obligado, no crece en cultivo puro, por lo que debe multiplicarse en las raíces de la planta hospedadora. La producción de inóculo por este método, además de ser cara y lenta para poder usarla con fines comerciales, tiene el peligro de contaminación por hongos patógenos que pueden desarrollarse a la vez que el hongo de la micorriza. Habrá que esperar, pues, al desarrollo de nuevos métodos de producción de inóculo.

Meteoritos basálticos

Las eucritas y las shergotitas, notablemente semejantes a rocas volcánicas de la Tierra y la Luna, proceden empero de algún otro lugar. Dan testimonio de la larga historia de la actividad volcánica registrada en el sistema solar

Harry Y. McSween, Jr., y Edward M. Stolper

Entre los miles de meteoritos que se han recogido y clasificado, una pequeña fracción, de unos 100, evidentemente cristalizó a partir de rocas fundidas en alguna parte del sistema solar hace unos 4500 millones de años, posiblemente no más tarde de 100 millones de años tras la formación del Sol y los planetas. ¿Cuál fue la fuente de calor capaz de fundir rocas en época tan temprana subsiguiente a la formación del sistema solar? ¿Y en qué clase de cuerpo tuvo lugar tal fusión? Una muestra todavía menor de meteoritos, en número de tres sólo, plantea un misterio aún mayor. Los tres objetos cristalizaron a partir de rocas fundidas hace no más de 1100 millones de años y presumiblemente fueron proyectados hacia el espacio en virtud de un impacto sobre un cuerpo muy similar en composición a la Tierra. Aunque se cree que la mayoría de los meteoritos proceden de cuerpos relativamente pequeños, esos tres muestran propiedades que indican que su origen fue un planeta grande, Marte quizá. Para explicar un origen tan improbable hay que invocar algún factor fuera de lo corriente, porque el impacto necesario para acelerar un fragmento de roca hasta la velocidad de escape, incluso desde un cuerpo tan pequeño como la Luna, es tan grande que nunca se han descubierto meteoritos de origen lunar.

Desde hace más de cien años se han conocido y estudiado meteoritos semejantes a rocas volcánicas terrestres. Estos proporcionaron pruebas de la existencia de procesos ígneos extraterrestres mucho tiempo antes del lanzamiento de misiones espaciales. Se sabe ahora que los procesos volcánicos dieron forma a la evolución de la mayoría de los cuerpos, grandes y pequeños, del sistema solar interno. Los "mares" de la Luna, de los que tomaron muestras los astronautas de las misiones Apolo, quedaron inundados de lava hace entre tres y cuatro mil millones de años. Las fotografías tomadas desde los vehículos espaciales Mariner y Viking revelan que los grandes volcanes de Marte son mucho mayores que las estructuras terrestres análogas, tales como las islas Hawai. El reciente descubrimiento por los Voyager de volcanes activos en Io, uno de los satélites de Júpiter, es el primer signo de la existencia del vulcanismo actual en un cuerpo distinto de la Tierra. Las observaciones de Mercurio realizadas por el vehículo espacial *Mariner 10* revelaron mares similares a los de la Luna, y algunas fotografías hechas por los vehículos rusos Venera que se posaron en Venus muestran rocas que parecen ser de origen volcánico. Sólo recientemente se ha apreciado en toda su amplitud la dimensión de la actividad volcánica en la Tierra al reconocerse que casi la totalidad de la corteza terrestre suboceánica está formada a partir de roca fundida que asciende por las dorsales centro-oceánicas.

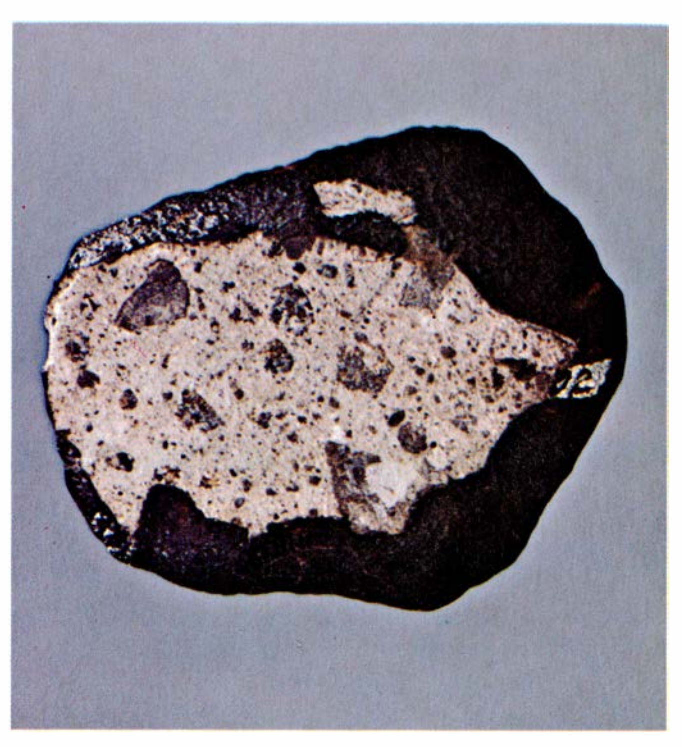
Las erupciones volcánicas constituyen la fase final de un proceso que comienza con la fusión de rocas en el interior del planeta, la "región fuente". La fuente de calor más corriente entre las que producen la fusión es la energía liberada por la desintegración de los elementos radiactivos. La materia en la región fuente, por lo general sólo parcialmente en estado de fusión, se halla formada por magma, o roca líquida, y cristales no fundidos. El líquido, más ligero, asciende gradualmente por encima de los cristales, más densos, por la acción de la gravedad, y se acumula en cámaras magmáticas. El proceso de segregación que impulsa el líquido hacia arriba puede durar desde cientos de años hasta cientos de millones, dependiendo sobre todo de la fuerza de la gravedad, que varía con la profundidad y el tamaño del planeta, de la naturaleza de los cristales en la región fuente, de la cantidad de líquido producido por el calor disponible y de la viscosidad del líquido.

Desde las cámaras de magma, la roca fundida busca su camino hasta la superficie a través de grietas o fracturas. En su camino hacia la superficie, la masa en fusión tiende a enfriarse y cristalizar parcialmente, de la misma manera que una taza de agua salada y caliente precipita gradualmente cristales de sal al enfriarse, cristales que crecen en el fondo y las paredes de la taza. Y de la misma manera que el agua en la taza se va haciendo menos salada a medida que la sal va precipitando, así cambia la composición del magma ascendente a medida que se enfría y los cristales se van separando de él. Aunque la analogía del agua salada constituye una simplificación excesiva, ilustra un importante punto: la mayoría de los magmas que acaban por hacer erupción han cambiado sustancialmente de composición a medida que se han ido enfriando en su camino hacia la superficie.

La fase final del proceso ígneo, la fase de erupción, es la única que se puede observar directamente. Es evidente que no todos los magmas producidos por fusión parcial en el interior de un planeta alcanzan la superficie. En muchos casos, los magmas quedan aprisionados y cristalizan completamente en cámaras debajo de la superficie.

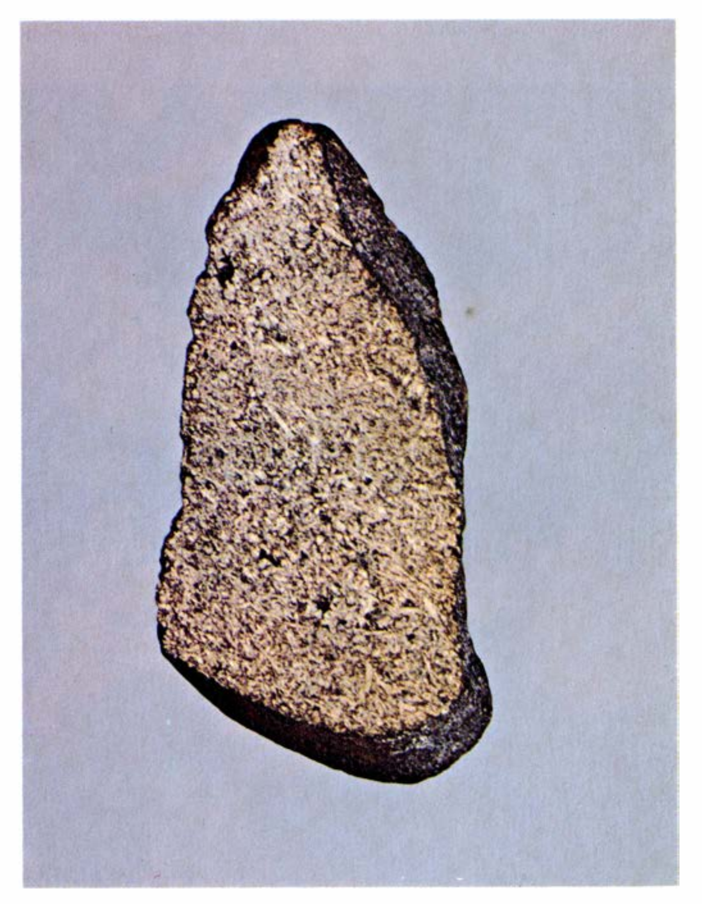
La simple observación de los ríos de lava y de los volcanes en la Tierra o en otro cuerpo celeste cualquiera revela gran cantidad de hechos sobre la historia de la evolución del planeta. En algún instante anterior a la erupción de lava, se debió suministrar suficiente calor para fundir parcialmente el interior del planeta. Si se supone que la fuente de calor es la desintegración de elementos radiactivos, pueden deducir-





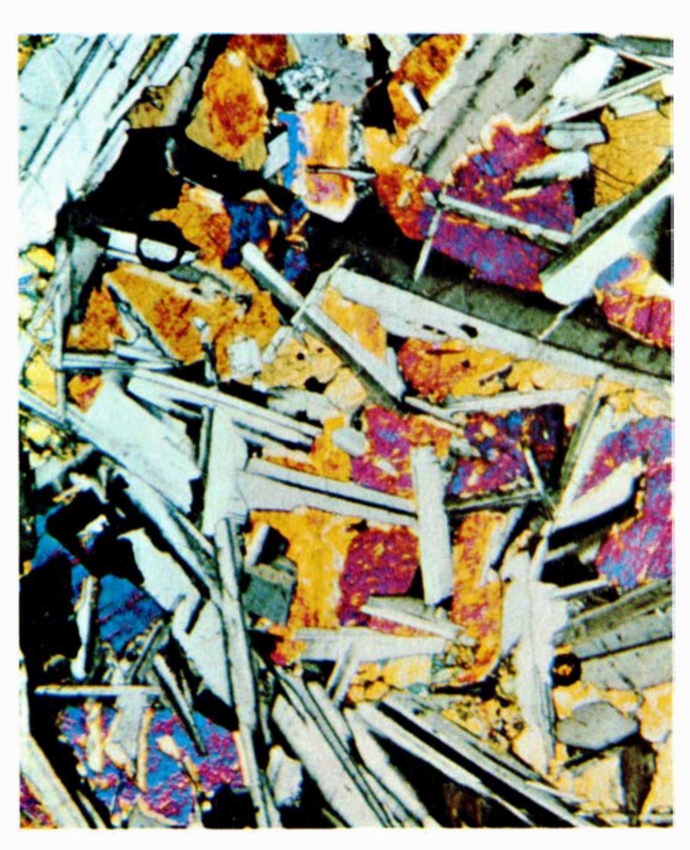
EUCRITA, perteneciente a una pequeña clase de meteoritos que cristalizaron a partir de rocas fundidas poco después de la formación del sistema solar, hace 4600 millones de años, mostrada en las dos fotografías. La de la izquierda muestra el oscuro y vítreo exterior del meteorito, lo que revela que éste cayó a gran velocidad a través de la atmósfera terrestre. La fotografía de la derecha muestra el corte del interior del meteorito. Tiene una textura brecciada, lo que indica que el meteorito consta de fragmentos que se soldaron en el cuerpo padre del cual fue expulsado. Esta eucrita particular es un fragmento de la lluvia de meteoritos de Pasamonte, que se observó en Nuevo México el 23 de marzo de 1933. Se muestra a la mitad de su tamaño.





SHERGOTITA, perteneciente a una clase todavía más pequeña de meteoritos. Se trata de un fragmento de lava que solidificó en algún lugar del sistema solar hace no más de 1100 millones de años. Sólo se conocen tres shergotitas; ésta es el meteorito original de Shergotty, cuya caída fue observada cerca de la ciudad de Shergotty (Shergati) en la India en 1865. La fotografía de la izquierda muestra el exterior del meteorito, la de la derecha el interior seccionado. Su interior está formado por cristales alargados del silicato piroxeno, que evidentemente se asentó en una orientación preferida a medida que la lava se enfriaba en una cámara de magma dentro del cuerpo padre del meteorito. Este ejemplar se ve también a la mitad de su tamaño natural. Tanto éste como el ejemplar de eucrita de las ilustraciones superiores fueron fotografiados en el Museo Nacional de Historia Natural de Washington.





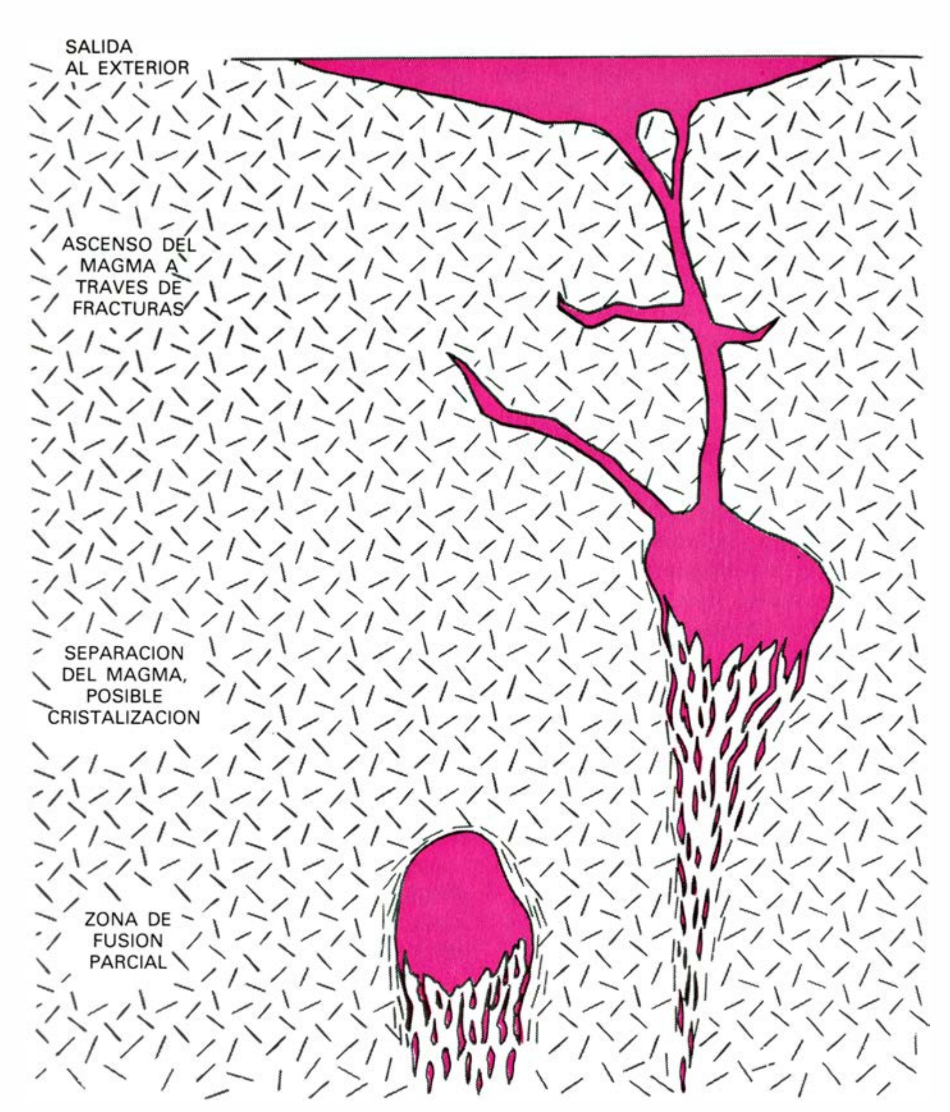
EUCRITAS Y BASALTO LUNAR. Aunque similares en mineralogía y textura, median suficientes diferencias entre ellos para indicar que las eucritas no proceden de la Luna. La micrografía, con luz polarizada, de la izquierda muestra el interior de una eucrita que cayó cerca de Juvinas (Francia) en 1821. El espécimen está formado por granos alargados de plagioclasa gris inmersos en cristales coloreados de pigeonita. En la micrografía de la derecha se aprecia el interior de un fragmento de basalto lunar. Este fragmento contiene también plagioclasa inmersa en pigeonita.





SHERGOTITAS Y BASALTO TERRESTRE. Se asemejan tanto en su composición y mineralogía que es difícil pensar que las shergotitas pudieran haberse originado en algún otro punto del sistema solar. No obstante, las pruebas de su origen extraterrestre son concluyentes. La micrografía de la izquierda muestra una shergotita que cayó en Zagami (Nigeria) en 1962. La micrografía de la derecha representa un basalto terrestre típico. Su diferencia fundamental estriba en que la shergotita contiene un vidrio denominado maskelinita, que se forma cuando la plagioclasa se ve sometida a una compresión intensa por impacto. La maskelinita aparece negra en la micrografía. Los cristales coloreados son augita y pigeonita. En la micrografía del basalto terrestre aparecen cristales alargados y grises de plagioclasa entre los granos coloreados de augita y pigeonita. No tiene maskelinita.

se muchas cosas acerca de la cantidad y distribución probables de tales elementos. Con independencia de la fuente de calor, es evidente que el líquido logró separarse de los cristales y hallar caminos para alcanzar la superficie. Por ejemplo, la aparente ausencia de actividad volcánica en la Luna desde hace unos tres mil millones de años no significa necesariamente que no haya habido fusión en el interior de la Luna desde aquella época; puede sólo indicar (como han sugerido Sean C. Solomon y sus colaboradores del MIT) que el enfriamiento y la contracción de la corteza de la Luna ha cerrado herméticamente todas las fracturas a través de las cuales podría haber ascendido el líquido.



ASCENSO DEL MAGMA, o roca líquida, a través de la corteza de un cuerpo planetario o asteroide, dibujado esquemáticamente. Las rocas de la región fuente se funden parcialmente a causa del calor generado por la desintegración de los elementos radiactivos. La parte fundida se separa de los cristales residuales, y por razón de densidad, emigra hacia la superficie. Parte de los líquidos quedan aprisionados en las cámaras de magma, donde recristalizan parcial o totalmente. Otras fracciones líquidas encuentran canales a través de los cuales alcanzan la superficie, se derraman y solidifican. Si la lava que cubre la superficie sufre el impacto de un cuerpo incidente, algunos fragmentos pueden ser lanzados al espacio y finalmente alcanzar la Tierra en forma de meteoritos. (Ilustración propiedad de Walken Graphics.)

Bajo condiciones favorables, se pueden deducir características de la mineralogía y de la composición química de la región fuente mediante el estudio de las lavas expulsadas. En otras palabras, se pueden usar las rocas volcánicas para explorar el interior de los planetas. Para ilustrar la manera como esto puede hacerse, considérese de nuevo una taza de agua, en la que esta vez haya en el fondo una capa de cristales blancos visibles. Si se prueba el agua y se la encuentra salada, es fácil deducir que los cristales son de sal. Si el agua es dulce, se deduce que los cristales son de azúcar. Si es al mismo tiempo salada y dulce, probablemente se llegaría a la conclusión de que son una mezcla de sal y azúcar. Lo mismo que los cristales del fondo de la taza dejan una marca distintiva en la química del agua, que se detecta mediante su sabor, así los cristales de las regiones fuentes de los interiores planetarios dejan una marca química distintiva en los líquidos que segregan las regiones fuente. Por tanto, si se pueden hallar e identificar rocas volcánicas cuya composición no haya cambiado grandemente por cristalización en el ascenso hasta la superficie, y el investigador es lo bastante hábil para reconocer la marca química de los cristales que quedaron atrás en la región fuente, éste puede hacer una hipótesis informada sobre la mineralogía y química del material padre en la región fuente. Debe añadirse que el paso crucial de identificar rocas volcánicas cuya composición no haya cambiado por cristalización en el camino desde la fuente hasta la supeficie es difícil y discutido. Al menos en la Tierra, tales rocas parecen ser raras; debe elegirse pues con gran cuidado las rocas a examinar para deducir los caracteres químicos de los cristales en la región fuente.

Con esta información básica, volvamos a las muestras de meteoritos y consideremos cómo pueden revelar la evolución térmica de los cuerpos de procedencia. La mayoría de los meteoritos no manifiestan indicaciones de origen ígneo. Una gran clase de meteoritos, las condritas, parecen ser mezclas de minerales inalterados que se condensaron a lo largo de un amplio intervalo de temperaturas antes de que los propios planetas se formasen a partir de la nube de gas y polvo conocida como nebulosa solar. Sin embargo, una pequeña fracción de meteoritos son de origen ígneo; se conocen como acondritas. El término se refiere a la ausencia en esas muestras de cóndrulos, pequeñas esférulas pétreas que se encuentran en las condritas. Las acondritas son las únicas muestras conocidas de rocas ígneas procedentes de fuera del sistema Tierra-Luna. Las rocas ígneas terrestres más antiguas conocidas se formaron hace unos 3800 millones de años; muchas de las acondritas cristalizaron entre hace 4400 y 4600 millones de años. Aunque hace más de 3800 millones de años ya se estaban formando rocas ígneas en la Tierra, ninguna de ellas ha sobrevivido porque todas se han fundido de nuevo. Se cree que la mayoría de las acondritas han sido arrancadas de sus cuerpos de origen por impacto de otros cuerpos, con diámetros comprendidos entre 500 kilómetros y quizás hasta un valor tan pequeño como 10 kilómetros, en el cinturón de asteroides situado entre las órbitas de Marte y Júpiter. Como veremos, incluso es posible identificar los cuerpos específicos de los que proceden algunas de las acondritas.

Uno de los grupos de acondritas mejor estudiados son las eucritas, del griego *eukritos,* que significa fácil de distinguir. Las eucritas están principalmente compuestas de cantidades casi iguales de los minerales denominados anortita y pigeonita. La anortita es una plagioclasa, un silicato de calcio y aluminio con la fórmula aproximada $CaAl_2Si_2O_8$. La pigeonita es un piroxeno de magnesio y hierro con la fórmula aproximada $MgFeSi_2O_6$; algunas muestras típicas incluyen también algo de calcio. Las eucritas son textural, mineralógica y químicamente semejantes a las rocas basálticas terrestres, que se definen de una manera algo vaga como rocas oscuras de grano fino que constan en gran parte de plagioclasas y piroxeno. Cabe pensar que no sería fácil reconocer las eucritas como meteoritos, ni siquiera por los geólogos, si no estuvieran visiblemente cubiertas por una corteza oscura y vítrea indicativa de que su parte exterior se ha fundido durante su rápido paso a través de la atmósfera terrestre.

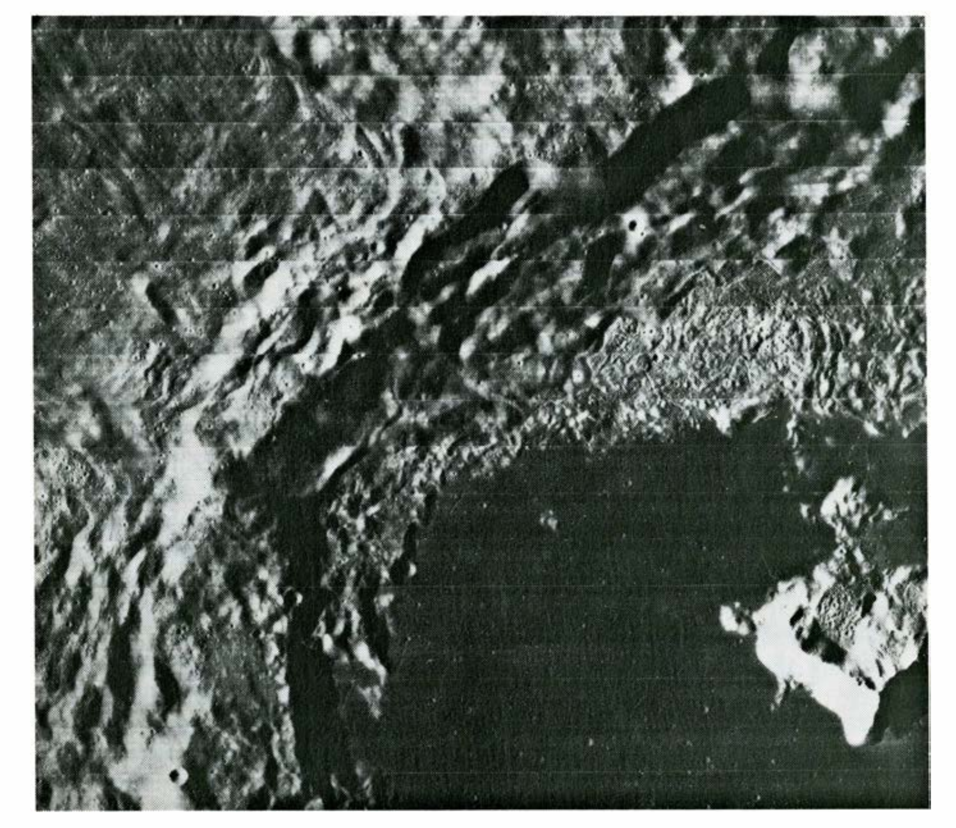
A pesar de la semejanza de las eucritas con los basaltos terrestres, se distinguen de ellos en varios respectos. El piroxeno dominante en los basaltos terrestres es la augita, un silicato rico en calcio que contiene diversas cantidades de magnesio y de hierro, mientras que la pigeonita, que es el piroxeno dominante en las eucritas, es pobre en calcio. Además, los basaltos terrestres son apreciablemente más ricos que las eucritas en elementos alcalinos (como el sodio, potasio y rubidio) y en otros elementos volátiles que se debieron condensar a partir de la nebulosa solar a temperaturas generalmente inferiores a unos 700 grados Celsius.

Es evidente que las eucritas se originaron en cuerpos planetarios formados predominantemente de materiales que se condensaron de la nebulosa solar a temperaturas más altas que aquellas a las que lo hicieron los materiales que formaron la Tierra. La relativa escasez de elementos volátiles en las eucritas aparece en la composición de la plagioclasa de las eucritas: es rica en calcio y pobre en sodio, mientras que las plagioclasas en los basaltos terrestres típicos vienen a tener en general la misma cantidad de ambos elementos.

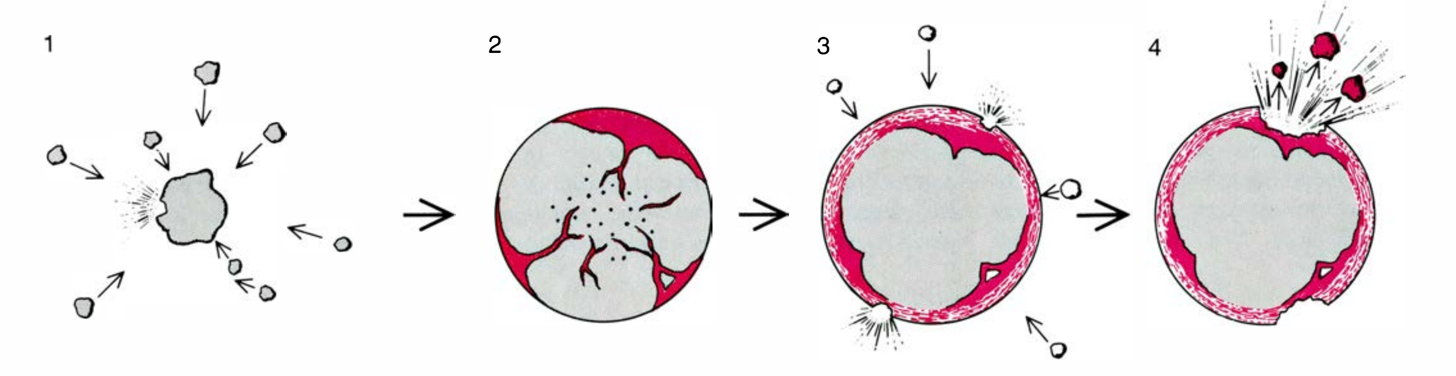
Los basaltos terrestres presentan, además, un estado de oxidación más fuerte que las eucritas. Cuando el metal hierro se expone a la acción del oxígeno, los dos elementos se combinan en diversas proporciones, según las condiciones; cuanto mayor es la relación de oxígeno a hierro, tanto más alto el grado de oxidación del compuesto. En las eucritas se encuentran granos de hierro metálico sin oxidar, además del óxido ferroso (FeO) de los silicatos minerales, mientras que en los basaltos terrestres el hierro no contenido en los silicatos se encuentra generalmente en forma de magnetita (Fe_3O_4). Además de esas diferencias de composición y estado de oxidación, las eucritas son generalmente brecciadas: constan de fragmentos angulosos que han quedado soldados entre sí. Se cree que esta estructura se debe al impacto de otros meteoritos sobre el material eucrítico cuando éste se hallaba todavía en la superficie del astro padre. Presumiblemente, este choque fue el que expulsó las eucritas de su superficie nativa, lanzándolas a órbitas que, en último término, cortaron a la de la Tierra.

Hay pruebas convincentes de que las eucritas cristalizaron a partir de un estado de fusión, bien en cámaras de magma o después de alcanzar la superficie del cuerpo padre. ¿Puede identificarse éste? ¿Hubo un solo cuerpo en el sistema solar que dio lugar a eucritas, o hubo varios de ellos? Aunque los basaltos lunares se parecen mucho a las eucritas, un estudio cuidadoso de las rocas lunares ha revelado diferencias que indican que las eucritas no pudieron haber venido de la Luna. De hecho, ningún meteorito conocido parece haberse originado en la Luna. Esto es de sumo interés, porque si no es posible expulsar rocas de la Luna y depositarlas en la Tierra a pesar del pequeño tamaño de la Luna, de su débil gravedad y de su proximidad, será difícil ver cómo los meteoritos pueden proceder de alguno de los planetas importantes, que son mayores que la Luna y están desde luego mucho más lejos de la Tierra. Por tanto, se supone generalmente que las eucritas son fragmentos arrancados de cuerpos pequeños del cinturón de asteroides, como lo son algunos otros tipos de meteoritos cuyas órbitas pudieron observarse antes de que chocaran con la Tierra.

En los últimos años, varios laboratorios han examinado intensivamente los espectros visible e infrarrojo de la luz solar reflejada de la superficie de otros planetas y asteroides. Se ha explorado ya la mayoría de los asteroides de diámetros superiores a 50 kilómetros. Los espectros de reflexión dependen de los minerales de la superficie del aste-



CRATER RELLENO DE LAVA típico de los "mares" lunares. La lava de tales cráteres escapó a través de fracturas en la corteza cuando la Luna fue bombardeada por cuerpos asteroidales en la primera época de su historia. Esta fotografía muestra parte del cráter de Tsiolkovsky, de 250 kilómetros de diámetro, en el lado opuesto de la Luna. Fue tomada en 1967 por el *Lunar Orbiter 3* durante un reconocimiento para seleccionar lugares de alunizaje para las misiones Apolo. Fue tomada a una altura de 1460 kilómetros por un sistema fotográfico que produce tiras de imágenes. El Norte lunar se halla en la parte superior.



LAS EUCRITAS PUDIERON FORMARSE en uno de los grandes asteroides, como Vesta; se cree que éste se constituyó por unión y soldadura de pequeños cuerpos hace 4600 millones de años (*1*). En los primeros 200.000 años, el calor, probablemente generado por elementos radiactivos de vida corta, fundió parcialmente el interior del cuerpo y parte del líquido llegó a la superficie en forma de lava (*2*). La superficie sufría un contínuo bombardeo de meteoritos (*3*). Los cráteres de impacto eyectaron al espacio fragmentos de la superficie de lava brecciada (*4*), alcanzando la Tierra alguno de ellos.

roide y de su composición. De aquí que, comparando los espectros de reflexión de asteroides con espectros de meteoritos analizados en el laboratorio, pueden hacerse deducciones sobre los materiales de la superficie del asteroide. En algunos casos, los espectros de los asteroides son casi exactamente iguales a los espectros de tipos conocidos de meteoritos. El único objeto del sistema solar con un espectro de reflexión similar al de las eucritas es Vesta, el segundo en tamaño de los asteroides, con un diámetro de unos 550 kilómetros. Aunque no hay acuerdo universal, la mayoría de los investigadores consideran probable que las eucritas sean fragmentos de Vesta, o al menos, que en la superficie de dicho asteroide existan lavas muy parecidas a aquellas a partir de las que cristalizaron las eucritas.

Si las eucritas vinieron de un cuerpo asteroidal, se debería poder deducir de ellas algunas cosas referentes a la mineralogía de la región fuente en el interior del cuerpo que dio lugar a los líquidos eucríticos o a la fusión parcial. ¿De qué manera debe estudiarse una roca volcánica para averiguar qué clase de cristales quedaron atrás en la región fuente? Volviendo a nuestra analogía anterior, supongamos, que en lugar de una taza agua con cristales blancos en el fondo, se nos entregara solamente una muestra del agua de la taza y se nos pidiera que identificásemos los cristales que quedaron atrás. Prescindiendo del procedimiento, un poco arriesgado, de probar el agua, ¿cómo lo haríamos?

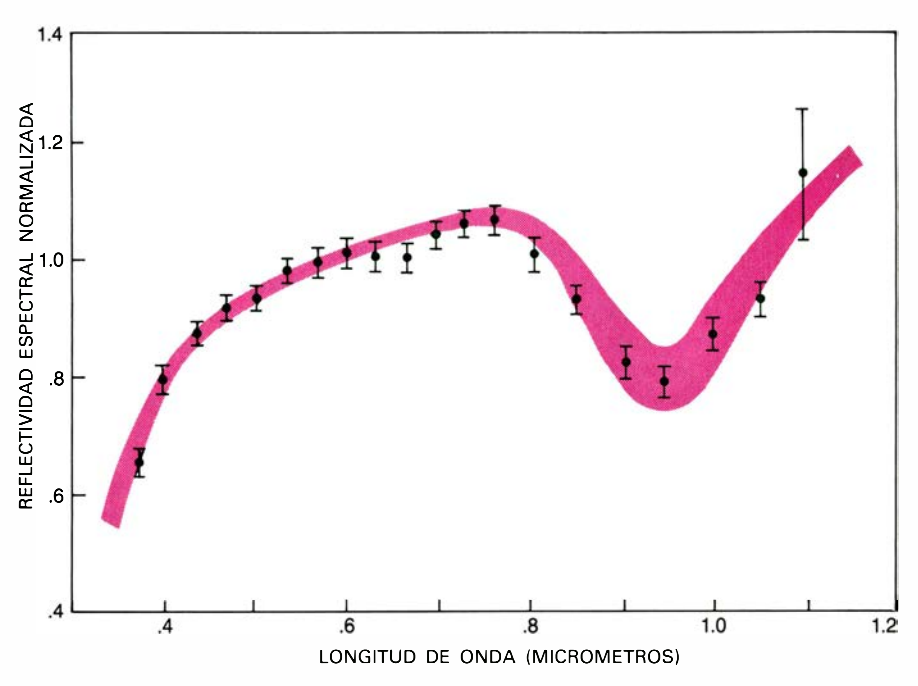
Una de las maneras es enfriar el agua lentamente a temperatura inferior a la de la habitación. Al descender la temperatura, se formarán nuevos cristales en el fondo y las paredes del recipiente a medida que la solubilidad de las sustancias disueltas decrece con la temperatura. Por su forma cristalina o bien por análisis químico podemos discernir si crecen cristales de sal o azúcar. El punto importante es que la identidad de los cristales desconocidos que quedaron en la taza se puede determinar identificando los cristales que crecen cuando el agua se enfría lentamente. De manera análoga podemos abordar el problema de identificar los cristales que quedaron atrás en las regiones fuentes de líquidos en el cuerpo padre de las eucritas. Se introduce una muestra de eucrita en un horno y se calienta hasta que se funda en su totalidad; entonces se la deja enfriar lentamente. El mineral o minerales que empiezan primero a cristalizar, es decir, aquellos con los cuales está en equilibrio el líquido eucrítico, deben tener la misma composición que los que quedaron en la región fuente.

Cuando se llevan a cabo tales experimentos con líquidos que poseen la composición de los meteoritos eucríticos, se encuentra que los líquidos están en equilibrio con tres tipos de cristales: olivino (un silicato de magnesio y hierro con la composición aproximada $MgFeSiO_4$), pigeonita y plagioclasa. La consecuencia clara es que esos tres minerales fueron los que quedaron atrás cuando los líquidos eucríticos se separaron de sus regiones fuentes en el interior del cuerpo padre de las eucritas.

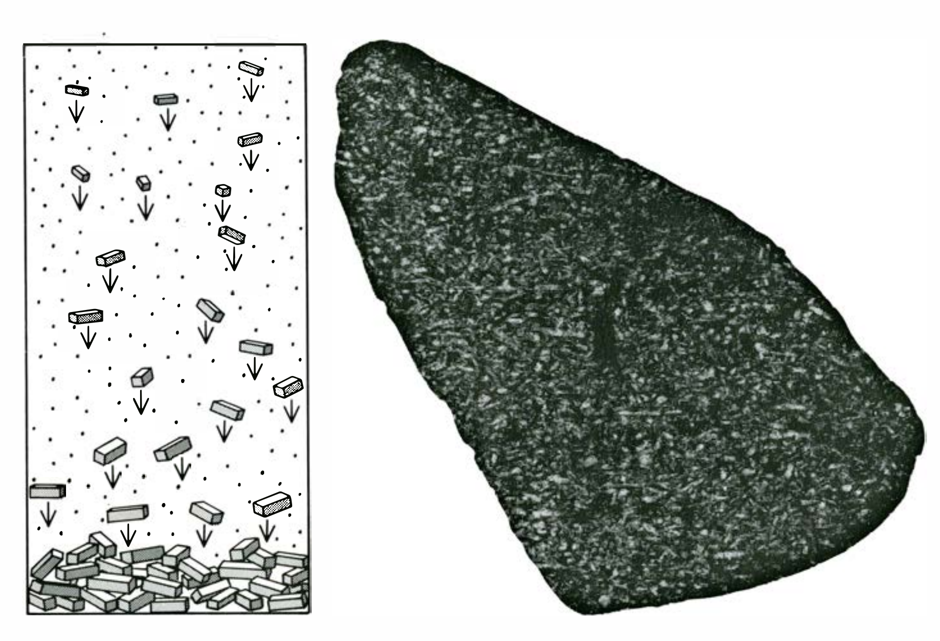
Hay varios problemas potenciales ligados a este método. Como hemos mencionado, es probable que la mayoría de las lavas se hayan enfriado y cristalizado parcialmente en el camino hasta la superficie, de modo que su composición original ha cambiado. Como resultado, los minerales que cristalizan primero cuando tales minerales se refunden en el laboratorio quizá no representen con exactitud los cristales que quedaron en la profundidad de las regiones fuentes. Un segundo problema es que la región fuente suele estar sometidas a una notable presión, ejercida por la masa de rocas que se encuentran encima de ella. Como a diferentes presiones son diferentes los minerales que cristalizan de un mismo líquido, se debe tratar de llevar a cabo los experimentos de cristalización a presiones que correspondan a la profundidad de las regiones fuente. Desgraciadamente, es difícil establecer la profundidad, incluso para las lavas terrestres. Sin embargo, si las eucritas proceden de cuerpos pequeños, tales como Vesta, la incertidumbre en la profundidad es una preocupación mínima; las presiones son bastante bajas, aun en el centro de tales cuerpos.

¿Qué indicaciones suministran tales experimentos de cristalización acerca de la composición en gran escala del cuerpo padre de las eucritas? Aunque los experimentos sugieren que el olivino, las pigeonitas y la plagioclasa fueron los que quedaron atrás, no especifican las proporciones de los tres minerales en la materia residual. El mismo líquido estaría en equilibrio tanto con un residuo que contuviese 99 por ciento de cristales de olivino como con otro que contuviese solamente el uno por ciento. Por fortuna hay otras marcas químicas en los líquidos que pueden ayudar a establecer los límites de las proporciones relativas de los minerales en el residuo. La concentración de elementos escasos, tales como las tierras raras lantano y europio, en las muestras de eucrita son especialmente útiles para este fin. Guy J. Consolmagno y Michael J. Drake, de la Universidad de Arizona, han llegado a la conclusión, a partir de datos de las tierras raras, de que los líquidos eucríticos podrían haber sido producidos por un 5 a 15 por ciento de fusión parcial de la peridotita, una roca compuesta principalmente de olivino con cantidades menores de piroxeno y plagioclasa.

Se cree también que la peridotita forma el grueso del manto superior de la Tierra, que es la región fuente de la mayoría de los basaltos terrestres. Como el manto de la Tierra está a presiones mucho más grandes que las del supuesto cuerpo padre de las eucritas, hay que esperar ciertas diferencias mineralógicas. Por ejemplo, en las peridotitas terrestres, la fase que contiene aluminio no toma generalmente la forma de plagioclasa, sino de granate o espinela, minerales que se originan a alta presión. De todas maneras, las peridotitas terrestres tienen mucho en común con las peridotitas que pudieran dar lugar a las eucritas. En ambas peridotitas, el olivino se presenta como mineral dominante, la fase principal de piroxeno es pobre en calcio y el principal mineral portador de aluminio es menos abundante que el olivino y el piroxeno.



SE SUGIERE EL ASTEROIDE VESTA como fuente de eucritas en virtud de la estrecha correspondencia entre el espectro de reflexión del asteroide (*barras negras*) y el espectro de reflexión típico de las eucritas (*banda en color*) medido en el laboratorio. En la actualidad se han medido los espectros de reflexión de virtualmente todos los varios cientos de asteroides catalogados con diámetro superior a 50 kilómetros. Solamente Vesta, con un diámetro de 550 kilómetros y el segundo en tamaño de los asteroides, tiene un espectro que se ajusta al de las eucritas. El mínimo en ambas curvas, a aproximadamente 0,95 micrometros, corresponde a una fuerte banda de absorción de la pigeonita, un silicato piróxeno.



ORIENTACION DE LOS CRISTALES DE PIROXENO en el meteorito de Shergotty. Da a entender que el material del meteorito cristalizó lentamente en una cámara de magma en el interior del cuerpo padre. El diagrama de la izquierda muestra esquemáticamente cómo los cristales alargados de piroxeno tienden a depositarse en el fondo de una cámara de magma en una orientación preferida. La fotografía del meteorito, a la derecha, se ha orientado de modo que sus cristales de piroxeno se alineen más o menos horizontalmente, como se supone que estaban cuando se separaron del magma hace mil millones de años.

Hay, sin embargo, diferencias químicas importantes entre los basaltos terrestres y las eucritas, que no pueden atribuirse a diferencias de presión entre las respectivas regiones fuente. El cuerpo padre de las eucritas es pobre en álcalis y otros elementos volátiles comparado con el manto superior terrestre. Además, éste contiene bastante piroxeno muy rico en calcio, y las regiones fuentes de las eucritas contienen poco o ninguno. Es probable que el manto superior de la Tierra esté también más intensamente oxidado que la fuente de las eucritas. En estos tres puntos, la región fuente de las eucritas se parece mucho más a la región fuente de los basaltos hallados en la Luna que a la de los basaltos terrestres.

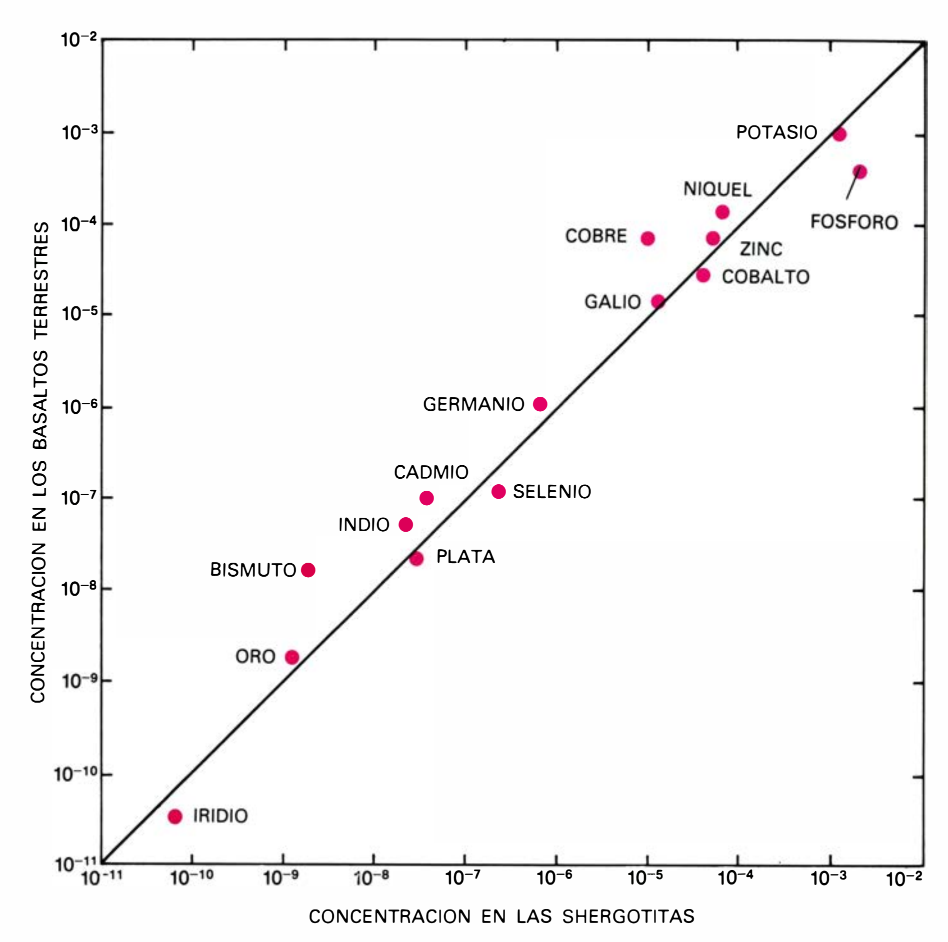
El rasgo más notable de las eucritas es su gran edad. Las edades de las rocas se pueden determinar midiendo con afinada precisión la cantidad de un isótopo radiactivo, tal como el rubidio 87, en la roca y comparando ese valor con la cantidad de algún isótopo estable que sea producto final de la desintegración del cuerpo radiactivo padre. En el caso del rubidio 87, que tiene una semivida de 47.000 millones de años, el producto final estable es el estroncio 87, elemento con un protón más y un neutrón menos. Debe hacerse una corrección, naturalmente, por la cantidad de estroncio 87 inicialmente presente en el magma. Una vez la roca cristaliza, no se puede añadir más rubidio, y el reloj radiométrico echa a andar. Tales determinaciones radiométricas establecen que todas las eucritas cristalizaron a partir de líquidos hace entre 4400 y 4600 millones de años, esto es, no más de 200 millones de años después de la formación del Sol y de los planetas. Las rocas más antiguas traídas hasta ahora de la superficie de la Luna revelan una antigüedad de unos 4200 millones de años; las más antiguas de la superficie de la Tierra son unos 400 millones de años más jóvenes. ¿Cómo se pudo generar el calor requerido para fundir las eucritas, unos 1150 grados C, en un cuerpo pequeño, en un período tan precoz de la historia del sistema solar?

El calor no pudo haber procedido de la lenta desintegración radiactiva de los isótopos de potasio, uranio y torio, que en último término generaron el calor necesario para fundir los interiores de la Tierra y la Luna. Se ha indicado, sin embargo, que cuerpos mucho más

pequeños que la Tierra y la Luna pudieron calentarse rápidamente por la desintegración del aluminio 26, un isótopo radiactivo de vida corta. La idea ha recibido apoyo como consecuencia del reciente trabajo de Gerald J. Wasserburg y sus colaboradores del Instituto de Tecnología de California, que han identificado un exceso de magnesio 26, el producto final de la desintegración del aluminio 26, en muestras del meteorito de Allende, una condrita que no ha sufrido fusión y por tanto retiene las características de la primitiva nebulosa solar. Se han señalado otras dos fuentes posibles de calor: el calentamiento del cuerpo padre de las eucritas por impacto de otros cuerpos y el calentamiento por el campo electromagnético del Sol juvenil.

Parece plausible el siguiente cuadro general de la evolución del cuerpo padre de las eucritas. Un cuerpo asteroidal, quizá Vesta, se formó hace 4600 millones de años. Mineralógicamente era una peridotita, con una composición química similar a la de los meteoritos condríticos. En este punto se parecía al manto de peridotita de la Luna más que al manto terrestre. Poco después de que se formase el asteroide, o tal vez durante su formación, su interior se fundió parcialmente, debido quizás al calor liberado por un isótopo radiactivo de vida corta. La roca fundida se separó de las regiones fuentes, dejando detrás como cristales residuales olivino, piroxeno pobre en calcio y plagioclasa. Los líquidos destinados a convertirse en eucritas emigraron hacia la superficie, algunos de ellos haciendo erupción en forma de ríos de lava. La actividad ígnea acabó en el intervalo de los 200 primeros millones de años inmediatos a la formación del cuerpo. Desde el mismo comienzo, el cuerpo asteroidal fue bombardeado por cuerpos de todos los tamaños. Los impactos produjeron el brecciado de la corteza de lava recién formada, y finalmente expulsaron fragmentos de la superficie, poniéndolos en una órbita que cortaba a la de la Tierra.

Un segundo grupo importante de acondritas son las shergotitas, nombre que procede de un meteorito que cayó cerca de la ciudad de Shergotty (o Shergatti), en el estado indio de Bihar, en 1865. Un segundo meteorito del mismo tipo cayó cerca de Zagami, Nigeria, en 1962. Un tercero (que completa el grupo conocido) se descubrió hace solamente tres años en un campo de hielo de la Antártida. Los tres cristalizaron a partir de roca fundida, al parecer hace no más de 1100 millones de años. Un rasgo no común de las shergotitas, que es lo que originariamente atrajo la atención hacia ellas, fue la naturaleza del vidrio de plagioclasa, que constituye casi la mitad de su masa. Este vidrio tiene la composición y apariencia de la plagioclasa, pero carece de la periodicidad atómica en gran escala que caracteriza los materiales cristalinos. Conocido como maskelinita, el vidrio se forma cuando se comprime la plagioclasa por impactos hasta presiones de 300.000 bar (unos 300.000 kilos por centímetro cuadrado). Recristaliza fácilmente a plagioclasa si se calienta hasta temperaturas de 400 grados Celsius durante unas horas. Las shergotitas presentan maskelinita y piroxenos intercalados en texturas típicas de rocas ígneas. Así pues, resulta claro que el choque que convirtió la plagioclasa en maskelinita



SEMEJANZA ENTRE SHERGOTITAS Y BASALTOS TERRESTRES que se extiende a los elementos menos abundantes y a los raros, sugiriendo que la región fuente de las shergotitas presenta un notable parecido con el manto superior de la Tierra. Los 15 elementos cuyas concentraciones se transcriben aquí difieren ampliamente en sus propiedades y comportamiento geoquímico. Por ejemplo, mientras el bismuto funde a 271 grados Celsius, el iridio lo hace a 2.350 grados. Al solidificar, el níquel se concentra preferentemente en fase metálica, el cadmio en fase de sulfuro y el germanio en fase de silicato. Los elementos que caen sobre la línea diagonal tienen la misma abundancia en ambos tipos de material.

MINERAL		METEORITO DE SHERGOTTY (PORCENTAJE)	SUELO MARCIANO (PORCENTAJE AJUSTADO)
DIOXIDO DE SILICIO	SiO_2	50,4	53,9
OXIDO FERROSO	FeO	19,3	19,7
OXIDO DE CALCIO	CaO	9,6	6,7
OXIDO DE MAGNESIO	MgO	9,3	10,0
OXIDO DE ALUMINIO	Al_2O_3	7,0	6,8
OXIDO DE TITANIO	TiO_2	9	1,0
OXIDO DE POTASIO	K_2O	2	1

COMPARACION DE LAS SHERGOTITAS CON EL SUELO MARCIANO: que nos revela sorprendentes semejanzas entre ellos, supuesto que la composición del suelo marciano se recalcula eliminando el azufre y el cloro. Parece probable que esos dos elementos representen "contaminación" por gases volcánicos en las muestras de suelo marciano analizadas por los vehículos Viking que se posaron en Marte. Las muestras contenían un 3 por ciento de azufre y un 1 por ciento de cloro. Estos elementos volátiles tienden a escapar y están en basaltos terrestres y lunares típicos en pequeñas cantidades.

tuvo lugar después que las shergotitas hubieran cristalizado a partir del estado fundido.

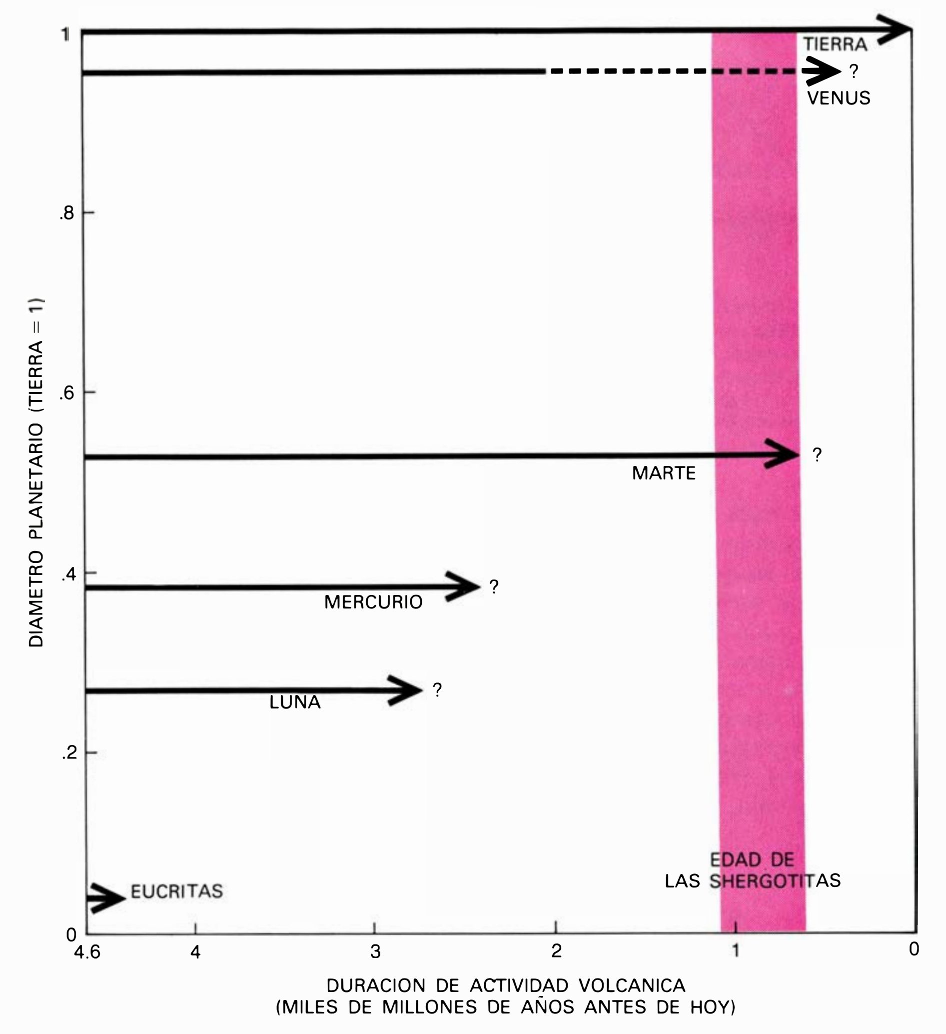
Si hacemos caso omiso de la presencia de maskelinitas, y a diferencia de las eucritas, las shergotitas podrían identificarse con las rocas terrestres basálticas típicas. Ambas contienen cantidades sustanciales de augita, además de pigeonita, y una y otra poseen magnetita; lo que prueba que las dos alcanzaron estados de oxidación más altos que las eucritas. La maskelinita en las shergotitas, lo mismo que la plagioclasa en los basaltos terrestres, contienen cantidades aproximadamente iguales de sodio y calcio. Esto contrasta con la plagioclasa en las eucritas, en las cuales la abundancia de esos elementos difiere de las halladas en los basaltos terrestres.

Las shergotitas difieren también de las eucritas en que la composición de los líquidos a partir de los cuales aquellas cristalizaron había sido alterada antes de que los líquidos se solidificasen en rocas. Tal cambio de composición, llamado diferenciación, puede tener lugar en los conductos a medida que los líquidos se enfrían en su camino hacia la superficie, o en las cámaras de magma, donde el enfriamiento ocurre con lentitud. Los cristales que se forman al principio y son más densos que el líquido pueden sedimentar poco a poco hasta depositarse en el fondo de la cámara de magma, apilándose en orientaciones privilegiadas. Tales acumulaciones de cristales de augita y pigeonita se encuentran en las shergotitas.

El meteorito original de Shergotty y el recientemente hallado en la Antártida han sido datados radiométricamente por Laurence E. Nyquist y sus colaboradores en el Centro Espacial Johnson de Houston, de la Administración Nacional de Aeronáutica y del Espacio (NASA), quienes aplicaron técnicas que habían sido desarrolladas para la datación de las rocas lunares. Determinaron que ambos meteoritos habían cristalizado a partir de líquidos hace entre 650 y 1100 millones de años. La incertidumbre en la edad proviene del suceso (choque) que creó la maskelinita y puso a cero algunos de los relojes radiométricos, pero las shergotitas se encuentran ciertamente entre las más jóvenes de las rocas extraterrestres conocidas.

Los espectros de reflexión de las shergotitas no ajustan con ninguno de los de los asteroides conocidos. Las shergotitas se acercan tanto a los basaltos terrestres que la composición química de sus regiones fuentes debe ser muy parecida a la del manto superior de la Tierra. El manto superior terrestre ha sufrido una evolución muy compleja. Se ha fundido parcialmente repetidas veces, produciendo los magmas que constituyen la corteza. Las concentraciones de elementos tales como níquel y cobalto en las regiones fuentes de los basaltos terrestres pudieron haberse establecido por interacción entre el manto de la Tierra y su núcleo metálico cuando éste se formó en la primera época de la historia de la Tierra, suceso que seguramente no se ha duplicado en detalle en ninguna otra parte del sistema solar. Sin embargo, el níquel y el cobalto se encuentran virtualmente en las mismas concentraciones en las shergotitas y en los basaltos terrestres. ¿Cómo es posible que otro cuerpo del sistema solar pudiera producir basaltos tan semejantes a los terrestres?

¿Cabe suponer que las shergotitas no fueran meteoritos sino trozos de basaltos terrestres? Ello explicaría su composición afín a la terrestre y su corta edad. Hay poderosas razones para dudar de esta posibilidad. Primero, las shergotitas india y nigeriana se vieron realmente caer, y tienen una corteza vítrea que se formó cuando pasaron a través de la atmósfera. Segundo, contienen indicación química, en la forma de isótopos radiactivos de vida corta, de que fueron bombardeados por rayos cósmicos en el espacio a lo largo de



DURACION DE LA ACTIVIDAD VOLCANICA en un planeta o planetoide relacionada con el tamaño del cuerpo. Venus, que es solamente un poco menor que la Tierra, está oculto por nubes, pero las técnicas de obtención de imágenes por radar han revelado la presencia en su superficie de estructuras semejantes a volcanes. Las fechas aproximadas del cese de la actividad volcánica en Marte y Mercurio se pueden estimar a partir del recuento del número de cráteres. Las rocas más modernas traídas de la Luna cristalizaron hace unos 2800 millones de años. Las eucritas proceden de un cuerpo padre, posiblemente Vesta, en el cual la roca fundida apareció no más de 200.000 años después de la formación del sistema solar, hace unos 4600 millones de años. Marte era al parecer todavía volcánicamente activo cuando las shergotitas cristalizaron a partir de roca fundida entre 650 y 1100 millones de años, y puede incluso ser activo hoy día. Aparte de la Tierra, el único cuerpo del sistema solar que tiene volcanes visiblemente activos es Io, uno de los satélites de Júpiter, según descubrimiento realizado hace un año por el vehículo espacial Voyager. Sólo un poco más grande que la Luna terrestre, Io parece deber su calor a las intensas mareas que se producen a causa de su proximidad a Júpiter. Las shergotitas son de origen volcánico.

EL MAYOR VOLCAN DE MARTE, el Monte Olimpo, aparece en una fotografía realizada en 1977 por el vehículo orbital *Viking 1*. Mucho mayor que ninguna estructura similar en la Tierra, el Monte Olimpo tiene 27 kilómetros de altura y unos 600 de diámetro en su base. Se formó por erupciones de lava de baja viscosidad. El Monte Olimpo pudo haberse mostrado activo hace no más de 500 millones de años e incluso serlo hoy día.



SUPERFICIE DE MARTE, fotografiada el 18 de mayo de 1979 por el *Viking 2*, vehículo que se posó en ella. Está sembrada de rocas que claramente solidificaron a partir del estado de fusión. El lugar es la Planicie Utopia, en el hemisferio Norte de Marte. Es por la mañana temprano y las rocas tienen una ligera capa de escarcha, que se cree ser una mezcla de agua y dióxido de carbono congelados. Tales rocas podrían ser similares a las shergotitas.

varios millones de años. Por tanto, si las shergotitas son rocas terrestres, hubieron de salir arrojadas de la Tierra por un impacto de considerable fuerza y colocadas en órbitas de "aparcamiento" capaces de retenerlas durante varios millones de años antes de permitirles su reingreso en la atmósfera terrestre. Aunque la expulsión por impacto explicaría la conversión de la plagioclasa en maskelinita, una explicación terrestre parece bastante traída por los pelos.

Quizá la objeción más seria contra el origen terrestre sea la existencia de pequeñas, pero significativas, peculiaridades químicas de las shergotitas. Por ejemplo, son más ricas en hierro y manganeso que los basaltos terrestres típicos, y, en este aspecto, se parecen a las eucritas. Además, los basaltos terrestren contienen cantidades pequeñas y mensurables de agua, de la que carecen en absoluto las shergotitas. Por último, Robert N. Clayton y sus colegas, de la Universidad de Chicago, han detectado pequeñas diferencias entre la composición isotópica del oxígeno en las shergotitas y la de los basaltos terrestres.

Si las shergotitas no se originaron en la Tierra, ¿dónde surgieron? Si se separaron, como las eucritas, de un pequeño cuerpo asteroidal, ¿cómo pudo tal cuerpo calentarse lo suficiente para que se haya producido fusión parcial en época tan reciente como 1100 millones de años y quizás hasta 650 millones de años?

Los isótopos radiactivos de vida corta, tales como los invocados para el calentamiento de un cuerpo del tamaño de Vesta hasta el punto en que se pudieran producir eucritas hace 4500 millones de años, se habrían desintegrado completamente hace 1000 millones de años. Por otro lado, los isótopos radiactivos de vida larga, tales como los que contribuyen el calentamiento de la Tierra hoy, no serían eficaces para calentar un cuerpo pequeño porque tal cuerpo no tendría capacidad de aislamiento suficiente para aprisionar el calor e impedir que se perdiera en el espacio. La relación entre el tamaño planetario y la capacidad de aislamiento proporciona una tosca correlación entre el tamaño de un cuerpo y la duración de su actividad volcánica. Las recientes misiones Voyager a Júpiter han revelado una importante excepción a esta correlación: Io, uno de los satélites de Júpiter que es sólo ligeramente mayor que la Tierra, es todavía volcánicamente activo. La fuente aparente de calor es el campo gravitatorio de Io o la interacción electromagnética con su gigantesco planeta. Para un cuerpo pequeño del tamaño de un asteoride, una fuente verosímil de calor capaz de fundir rocas en un intervalo reciente, del orden de 1000 millones de años, podría ser el choque con otro cuerpo. No es fácil demostrar, sin embargo, que tal choque llegara a originar un producto fundido que tuviera la composición de una shergotita.

De todas maneras, si las shergotitas son realmente fragmentos de un asteroide, ello significa que hay un pequeño cuerpo en el sistema solar con características químicas que se había creído eran exclusivas de la Tierra o al menos de un planeta de tamaño comparable. Independientemente del tamaño del cuerpo padre de las shergotitas, está ahora claro que muchas de esas características no son exclusivas de la Tierra. Si el cuerpo padre de las shergotitas fuera de tamaño asteroidal, ello podría significar que el cuerpo padre de las shergotitas y el manto superior de la Tierra se formaron de materiales semejantes en la fase temprana de la historia del sistema solar. En este caso, muchas de las características químicas de las rocas terrestres que antes se creía reflejaban únicamente la complicada evolución de la Tierra habrían sido en realidad heredadas de los materiales de construcción a partir de los cuales se aglomeró nuestro planeta. Parece pues verosímil que el cuerpo padre de las shergotitas sea un trozo de material de construcción desechado del tipo del cual se formó el manto superior terrestre.

Si ni la Tierra ni un cuerpo asteroidal se consideran fuentes verosímiles de las shergotitas ¿qué alternativas quedan? Creemos poder prescindir del todavía activo satélite de Júpiter, Io, por dos motivos. Mediciones recientes sugieren que la superficie de Io abunda tanto en azufre y sodio que la composición de sus productos volcánicos sería totalmente diferente de la de las shergotitas. La segunda objeción es que cualesquiera fragmentos de Io que fueran proyectados por el impacto de otros cuerpos tendrían muy poca probabilidad de escapar de la inmensa atracción gravitatoria de Júpiter.

La única fuente alternativa de las shergotitas que encontramos atractiva es Marte. Las fotografías de los Mariner y Viking han revelado la existencia de volcanes gigantescos en su superficie. A partir del pequeño número de cráteres de impacto en algunos ríos de lava marcianos, se puede estimar que el planeta mostró, seguramente, actividad volcánica hasta época tan reciente como hace 500 millones de años y quizá sera activo incluso hoy día. La gran objeción a Marte como fuente de las shergotitas es la total ausencia de meteoritos procedentes de la Luna. Un impacto capaz de expulsar un fragmento de la superficie marciana lanzándolo a una órbita que corte a la de la Tierra parece que debe ser todavía menos probable que tal suceso en la Luna, en vista del menor tamaño y mayor proximidad a la Tierra de esta última.

El caso marciano no queda descartado, sin embargo. George W. Wetherill, del Departamento de Magnetismo Terrestre de la Institución Carnegie de Washington, sugirió recientemente que los hielos interpuestos en el suelo y almacenados bajo la superficie de Marte podrían alterar las características de impacto del planeta. Si los hielos fuesen rápidamente vaporizados por un objeto incidente, los gases en expansión podrían ayudar a los fragmentos expulsados a alcanzar la velocidad de escape.

Hay un último e intrigante indicio que puede ligar las shergotitas a Marte. Los dos vehículos espaciales Viking que se posaron en Marte hicieron análisis cuantitativos del suelo del planeta en dos puntos y lo encontraron bastante rico en azufre y cloro. Esto se ha interpretado generalmente en el sentido de que las muestras de suelo examinadas no representan rocas ígneas jóvenes sino muestras que han sido alteradas por la intemperie. Si se supone que el azufre y el cloro son en realidad agentes de la intemperie y se resta esos dos elementos de los análisis del suelo, se encuentra que la composición química del suelo marciano se parece mucho a la composición de las shergotitas. Sólo en óxido de calcio discrepan entre sí las dos muestras en más del 10 por cien [*véase la ilustración inferior de la página 25*].

Tal vez haya de transcurrir mucho tiempo antes de que los geólogos tengan en sus manos una muestra inequívocamente identificada de suelo o roca marciana, y todavía más antes de que puedan examinar una muestra inequívocamente identificada de Vesta o de otro de los grandes asteroides. Mientras tanto, continuarán cayendo meteoritos y desafiando a los geólogos y geoquímicos a que adivinen su origen. La búsqueda de las fuentes de las shergotitas y eucritas ilustra la clase de trabajo policial que se necesita realizar.

Chaquete por ordenador

El chaquete es terreno apto para ensayar ciertos principios de inteligencia artificial. El programa BKG 9.8 ha sido el primero en derrotar al campeón mundial Villa

Hans Berliner

En julio del año pasado, un programa mío diseñado para jugar al chaquete (en inglés, backgammon), que bauticé BKG 9.8, consiguió derrotar al campeón del mundo de la especialidad, el italiano Luigi Villa, en encuentro celebrado en Monte Carlo, y dotado de 5000 dólares para el vencedor. Fue la primera vez que un programa de ordenador conseguía ganarle a un campeón mundial, ya fuera de juegos de cartas o de juegos de tablero. Paul Magriel, el anterior campeón, quien aportó buena parte del saber y experiencia necesarios para poner a punto el programa BKG 9.8, estimaba antes del encuentro que el programa tendría de un 35 a un 40 por ciento de probabilidades de vencer. Tal opinión me pareció de un optimismo desbordante. Pero a la hora de la verdad, el programa no sólo supo aprovechar al máximo las puntuaciones de los dados, sino que ejecutó un juego preciso e imaginativo, haciéndose claramente con la victoria. Si, a diferencia de los mejores programas de ajedrez disponibles, el BKG 9.8 consigue tan brillantes resultados es porque opera más por evaluación posicional que a base de puros cálculos —de fuerza bruta, podríamos decir— y su estilo de juego es muy semejante al de los expertos humanos.

Mi trabajo sobre chaquete por ordenador forma parte de una gran rama de las ciencias de computación, llamada inteligencia artificial, cuyo objetivo es preparar programas capaces de realizar cosas que, de ser ejecutadas por humanos, serían calificadas como inteligentes. Los investigadores de este campo se han interesado siempre por los juegos, porque para triunfar en ellos se requiere indudable inteligencia. En los juegos, las reglas están definidas con precisión, y, por lo general, se dispone de criterios para valorar la actuación de los jugadores. Por ello, con frecuencia es factible enfrentar directamente programas y contrincantes humanos bajo condiciones controladas. Podemos entonces medir con precisión la capacidad de juego del programa (y, por tanto, su inteligencia.)

Durante seis años estuve trabajando en un programa de ajedrez del que confiaba pudiera afrontar el juego como lo haría un gran maestro. Por mi parte, me encontraba razonablemente cualificado para ello. He sido a lo largo de 15 años uno de los 12 jugadores punteros de los Estados Unidos; en 1968 gané el campeonato del mundo de ajedrez por correspondencia. Hay varias clases de programas de ajedrez; una de ellas está formada por los llamados programas de exploración, diseñados para que examinen hasta profundidad prefijada todas las posibilidades. En el transcurso de mi trabajo sobre programas de ajedrez tropecé con varias categorías de problemas metodológicos que presentaban grandes dificultades al análisis por exploración. Uno de ellos es el llamado efecto de horizonte, que se presenta cuando el programa explora hasta profundidad suficiente como para descubrir que el juego va a presentar alguna dificultad, y después, sin embargo, actúa como si la dificultad no existiera. En realidad, el programa detecta la dificultad, pero la pospone porque ha descubierto un rodeo que contiene, pongamos por caso, un par de jugadas intrascendentes pero obligatorias en esa coyuntura. Como todas las variantes son analizadas hasta la misma profundidad, en la nueva variante la búsqueda concluye antes de que el problema se presente. El resultado es que el programa actúa como si la dificultad hubiese sido definitivamente eliminada. De haber profundizado un poco más por esa rama se hubiera tropezado con el mismo problema.

Un jugador avezado nunca hubiera cometido un error de tanto bulto, lo que indica que los jugadores humanos enfocan el juego desde otra perspectiva. Para ellos, las jugadas forman parte de acontecimientos, y en tanto quede algún acontecimiento importante por explorar en una rama cualquiera, el jugador humano porfía en su búsqueda. Al fijarse no en los acontecimientos, sino en las jugadas, los programas de hoy juegan torpemente en situaciones claras.

Por otra parte, infundir en un programa elementos de análisis posicional resultaba extraordinariamente difícil. Los maestros de ajedrez pueden estar de acuerdo en que un determinado ataque valga, pongamos por caso, un par de peones. En apreciaciones como ésta entran en juego gran cantidad de diversas informaciones, siendo ya muy notable el hecho de que puedan traducirse en valoraciones cuantitativas. Casi todo el mundo pondría en entredicho que un programa pueda realizar semejantes proezas. Me parecía que la introducción de normas que permitieran a los programas juicios así iba a requerir un número increíblemente grande de reglas, número seguramente prohibitivo, que especificasen en qué condiciones ciertas características de una posición son ventajosas, y en qué medida lo son. En consecuencia decidí estudiar cómo podrían combinarse variables asociadas a rasgos importantes de una posición y crear con ellas funciones aritméticas que pudieran responder a ligeras diferencias de matiz. El ajedrez no resultaba adecuado para una investigación de esta naturaleza, porque gran parte del proceso de enjuiciamiento implica ensayar las alternativas para ver qué características producen fruto en determinadas posiciones. Lo que yo necesitaba era un dominio donde fuese posible comparar dos situaciones y formular juicios sobre cuál es preferible, sin tener que ocuparse de analizarlas exhaustivamente, como exigen los problemas ajedrecísticos. Cuando llegué a familiarizarme con el chaquete me di cuenta de que reunía las propiedades deseadas, y me

propuse poner a punto un programa para este juego.

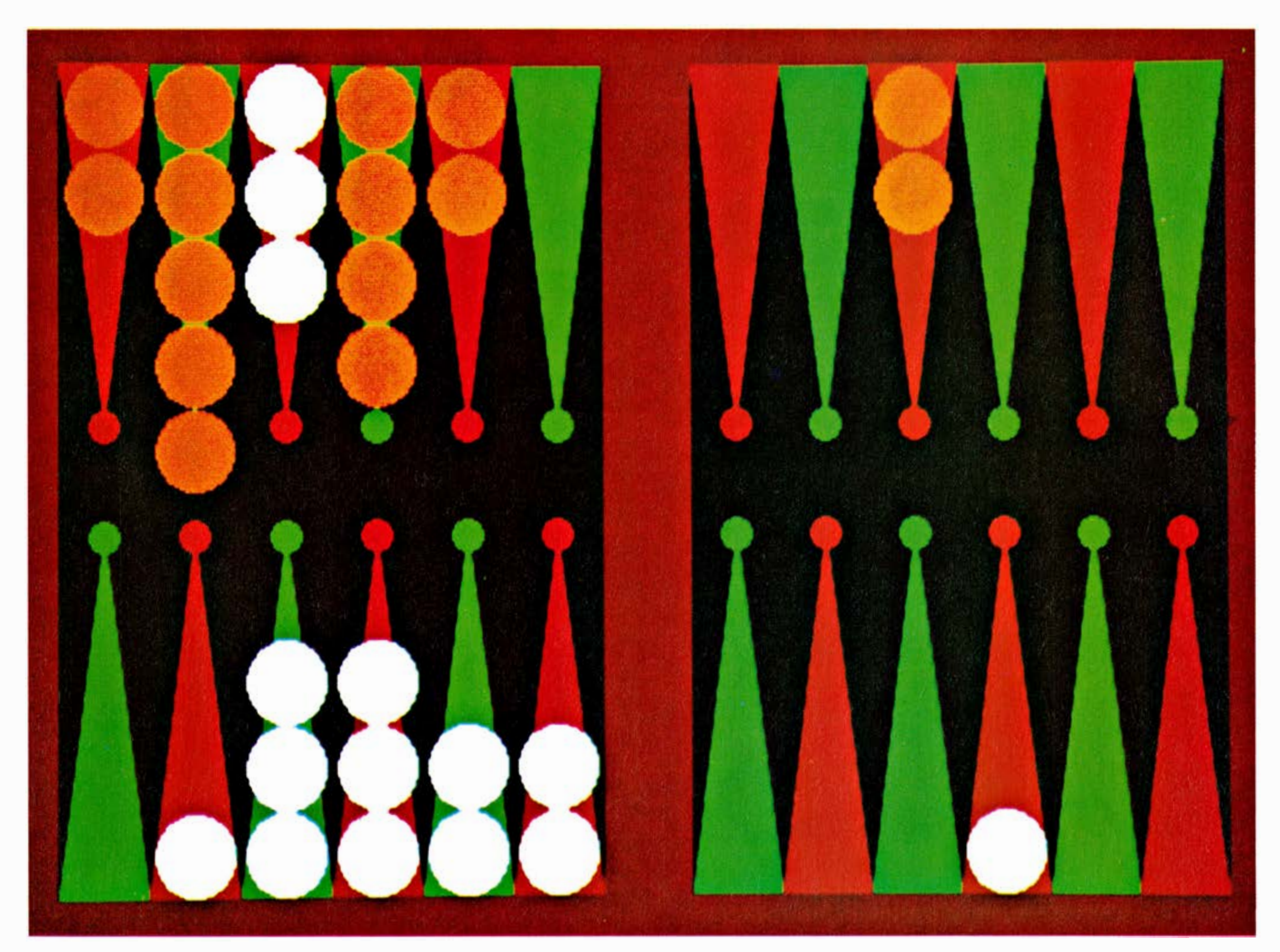
Describiré sucintamente las reglas del chaquete, por si alguno de los lectores no estuviera muy al tanto. El chaquete es un juego de dados. Dos jugadores tratarán de retirar del tablero la totalidad de sus 15 peones antes de que el oponente lo consiga con los suyos. El tablero consta de 24 "puntas" triangulares, distribuidas en cuatro cuadrantes o cuarteles: las "mesas" interior y exterior de cada jugador. Vemos en la página siguiente la posición inicial de las piezas. Las piezas han de ir desplazándose de una punta a otra, conforme al número de puntos que obtenga el jugador de turno al lanzar dos dados.

Cuando las puntuaciones de los dados sean distintas, el jugador puede servirse de cada una de las puntuaciones para mover una pieza distinta, o bien servirse de la puntuación suma para mover sólo una pieza. Aun así, cada puntuación se toma individualmente, y cuando las usa el jugador para mover sólo una pieza, no hace un movimiento, sino dos. Por ejemplo, si uno de los dados muestra un seis y el otro un dos, es lícito mover una sola pieza, primero seis puntas, y seguidamente, otras dos (o primero dos, y luego seis, si conviene). En estas jugadas de dos pasos, cuando una pieza ocupa momentáneamente una punta entre ambos se dice que "toca" en ella.

Si al lanzar los dados ambos sacan el mismo número, el jugador utiliza ese número cuatro veces. Por ejemplo, si han salido dos "doses" es lícito (1) mover una pieza un total de ocho puntas, (2) mover dos piezas, cuatro puntas cada una, (3) dos piezas a dos puntas cada una, más otra pieza, cuatro puntas, (4) una pieza seis puntas, y otra dos, y (5) cuatro piezas, a dos puntas cada una.

Cuando al menos dos piezas del mismo bando ocupan una punta, se dice que está "hecha" o completa, y ninguna pieza del contrario puede tocar ni quedarse en ella. Las puntas completas son muy ventajosas, pues no sólo bloquean el avance del contrario sino que sirven de base segura para las propias. Las piezas solitarias en una punta ("blots") están en precario, pues son "alcanzadas" o "heridas" cuando una pieza del contrario toca o invade la punta que ocupan, causando "baja" y saliendo del tablero. Los peones de baja descansan en la barra central que separa los cuarteles interiores de los exteriores, y antes de mover ninguna otra pieza, han de ser reingresados al principio del tablero.

Cuando un jugador consigue conducir todas sus 15 piezas hasta su mesa interior, puede empezar a retirarlas del tablero, de acuerdo con las puntuaciones que muestren los dados. El jugador que primero consiga retirar del tablero todas sus piezas es vence-

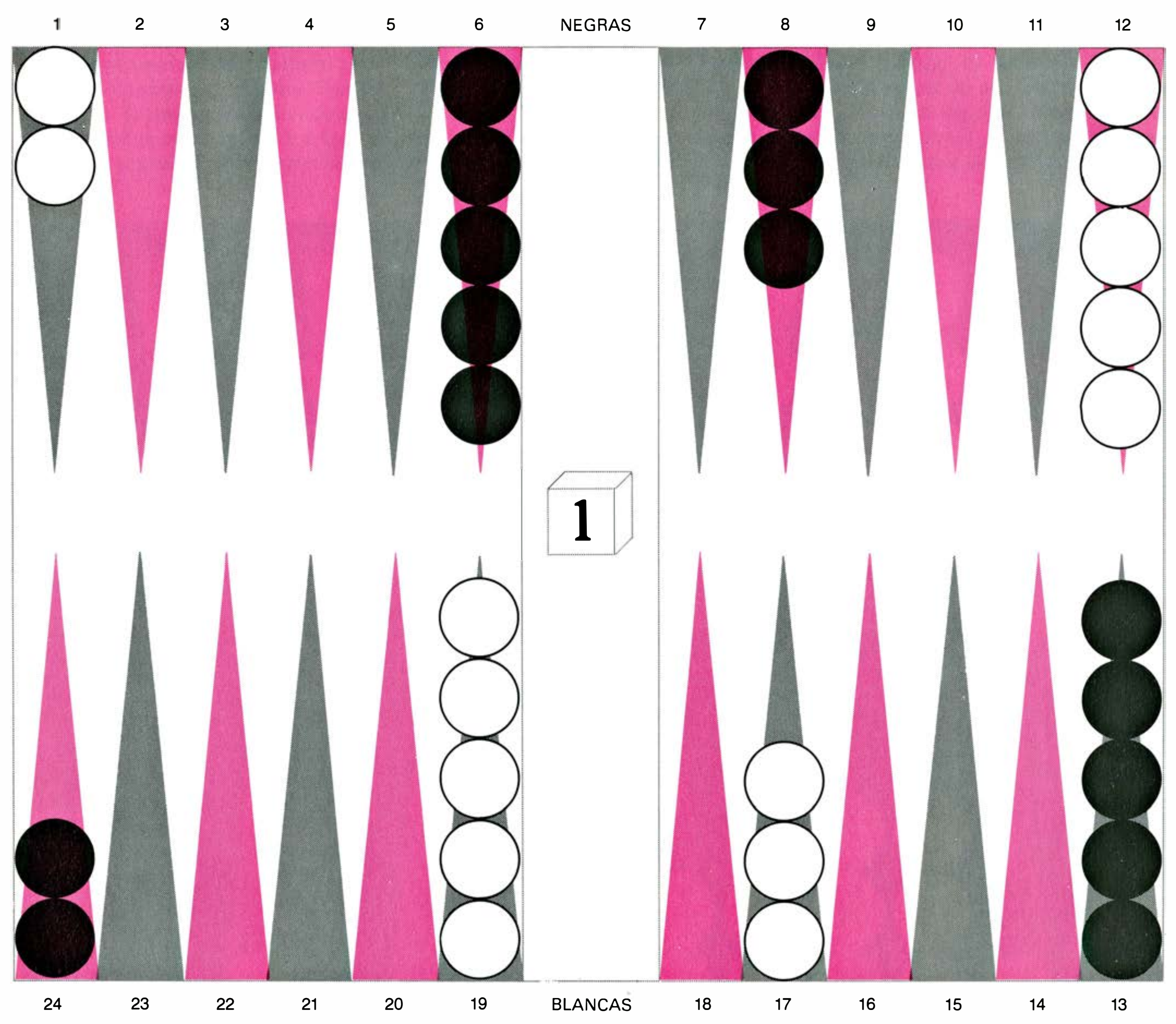


PRESENTACION EN PANTALLA de una situación crítica de la primera partida del encuentro entre el programa BKG 9.8, preparado por el autor, y el campeón del mundo de chaquete, el italiano Luigi Villa. El vencedor recibiría un premio de 5000 dólares. El programa jugaba con las piezas anaranjadas, y movía en sentido antihorario; Villa llevaba las blancas y avanzaba en sentido horario. Para ganar la partida el jugador debe comenzar por llevar todas sus piezas hasta las seis "puntas" o triángulos de su cuartel de llegada, y a continuación, retirarlas del tablero antes de que lo consiga su adversario. El cuartel de llegada del programa está arriba, a la izquierda; el de Villa, abajo a la izquierda. En la posición que vemos el programa lleva la delantera, porque el campeón mundial tendría todavía que mover las tres piezas blancas situadas arriba a la izquierda 16 puntas antes de tenerlas en su cuartel de llegada. La posición se analiza en la ilustración superior de la página 35. (Visualización dada por Three Rivers Computer Corporation.)

dor. Hay victorias de tres tipos: normales, cuyo valor es la puesta estipulada, cuando el perdedor ha conseguido sacar del tablero al menos una de sus piezas; la victoria doble o chaquete (en inglés, "gammon"), de valor doble de la puesta acordada, cuando el perdedor no ha conseguido sacar del tablero ninguno de sus peones; y la triple ("backgammon"), cuando el perdedor no sólo no ha conseguido retirar del tablero ninguna de sus piezas, sino que además tiene todavía alguna retenida en la mesa interior de su contrario, o de baja, en la barra central.

Como juego de apuestas, casi todo el interés del chaquete se debe al llamado cubo de duplicar. En cualquier momento, antes de que el jugador de turno lance sus dados puede proponerle al contrario duplicar la apuesta, volviendo para ello el cubo, con la cara marcada con 2 hacia arriba. Su oponente puede rehusar la oferta, en cuyo caso entrega la partida y pierde la puesta, o aceptarla, y entonces la partida se juega por una puesta doble. El jugador que acepta el envite adquiere control del cubo de duplicar, lo que quiere decir que sólo él podrá ofertar la segunda duplicación de la puesta. Entre jugadores avezados el número de duplicaciones suele ser pequeño, y pocas veces marca el cubo más de 4.

Tal es la naturaleza del juego que me propuse programar. La pieza clave de los programas de juego es un generador de jugadas que, aceptando como entrada una posición de la partida, va produciendo todas las jugadas lícitas (y solamente las jugadas lícitas) que pueden realizarse desde dicha posición. A continuación el programa decide qué jugadas va a examinar con mayor profundidad. Un programa de ajedrez típico investiga todas las jugadas lícitas que es factible realizar a partir de cada posición dada. Las posibles continuaciones del juego se organizan en un diagrama arbóreo. Se examina cada rama del árbol hasta que el programa



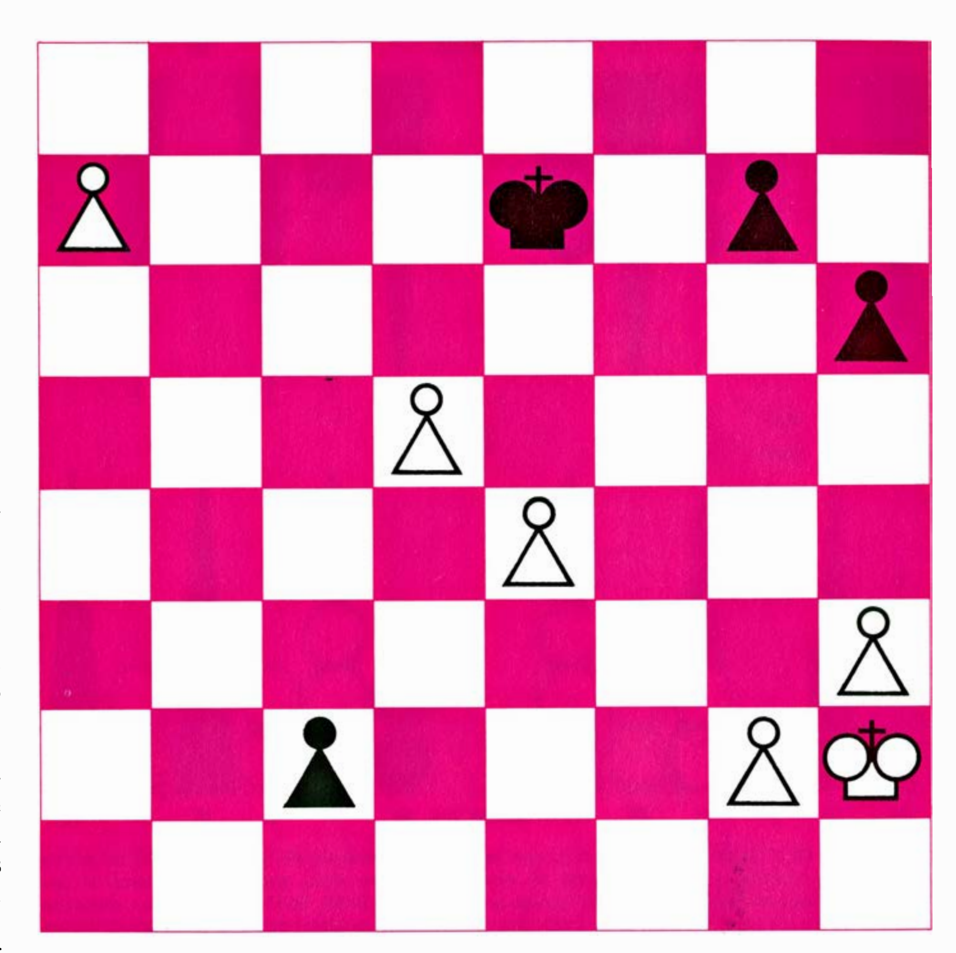
POSICION INICIAL de una partida de chaquete. Las piezas están dispuestas simétricamente. El tablero está formado por 24 puntas rojas y negras, repartidas en cuatro cuadrantes: la "mesa" interior de las negras (puntas números 1 a 6), la mesa exterior de las negras (las puntas números 7 a 12), la mesa exterior de las blancas (formada por las puntas 13 a 18) y la mesa interior de las blancas (puntas 19 a 24). Las piezas se mueven de unas puntas a otras, conforme a las puntuaciones obtenidas al lanzar dados. Las piezas blancas se desplazan en el sentido de las agujas del reloj, y por tanto van recorriendo las puntas en orden numérico creciente; las negras lo hacen en sentido antihorario, y recorren las puntas en orden decreciente. Es fácil comprender la notación utilizada para designar jugadas. Por ejemplo, 7-10 significa que una pieza ha sido movida de la punta número 7 a la número 10. El cubo situado en la barra divisoria entre mesas interiores y exteriores se llama cubo de doblar. El 1 indica que se está jugando con la puesta inicial. En cualquier momento, el jugador de turno puede, antes de lanzar los dados, ofrecer la duplicación de la puesta, volviendo el cubo a la cara marcada 2. Su contrario tiene que aceptar la oferta o rechazarla. De aceptarla, el juego se prosigue, pero ahora la apuesta es doble de la inicial. Si la rehúsa, el juego termina con victoria del jugador de vez, y el otro pagará la puesta. Cuando un jugador acepta un "doble" sólo él puede ofrecer la duplicación siguiente.

encuentra alguna razón para poner fin a la búsqueda. Seguidamente, el programa aplica una función de evaluación a la posición terminal a fin de llegar a un valor cuantitativo que exprese qué jugador se encuentra en mejor situación, y cuánto mejor. El programa elige su jugada comparando este valor con los correspondientes a las restantes ramas.

Por lo general, en los programas de ajedrez el examen de cada rama se realiza hasta una profundidad máxima prefijada. En ajedrez, el coeficiente de ramificación (número medio de jugadas realizables a partir de cada nudo del diagrama de juego) está alrededor de 35, y un programa rápido puede examinar hasta la profundidad de seis jugadas en los tres minutos por jugada que se conceden en los torneos. En cambio, el coeficiente de ramificación del chaquete es superior a 400, pues los dados muestran 21 posibles resultados distintos y hay alrededor de 20 maneras distintas de utilizar cada posible puntuación. Tan alto valor del coeficiente muestra que la exploración exhaustiva no podrá servir para diseñar programas de ordenador capaces de jugar con éxito al chaquete.

Lo que hace mi programa es generar todos los movimientos válidos en cada posición del juego, y evaluarlos directamente, sin exploración previa. El programa selecciona la jugada con máxima evaluación y la ejecuta en el juego. La función de evaluación toma en cuenta factores como seguridad de las piezas, medida en que se encuentran bloqueadas, grado de ventaja de cada bando en la carrera para retirar piezas, y otras semejantes. También el cubo de duplicar tiene asignada una función de evaluación, que decide si debe ofrecerse un doble, o si deben aceptarse las ofertas del contrario.

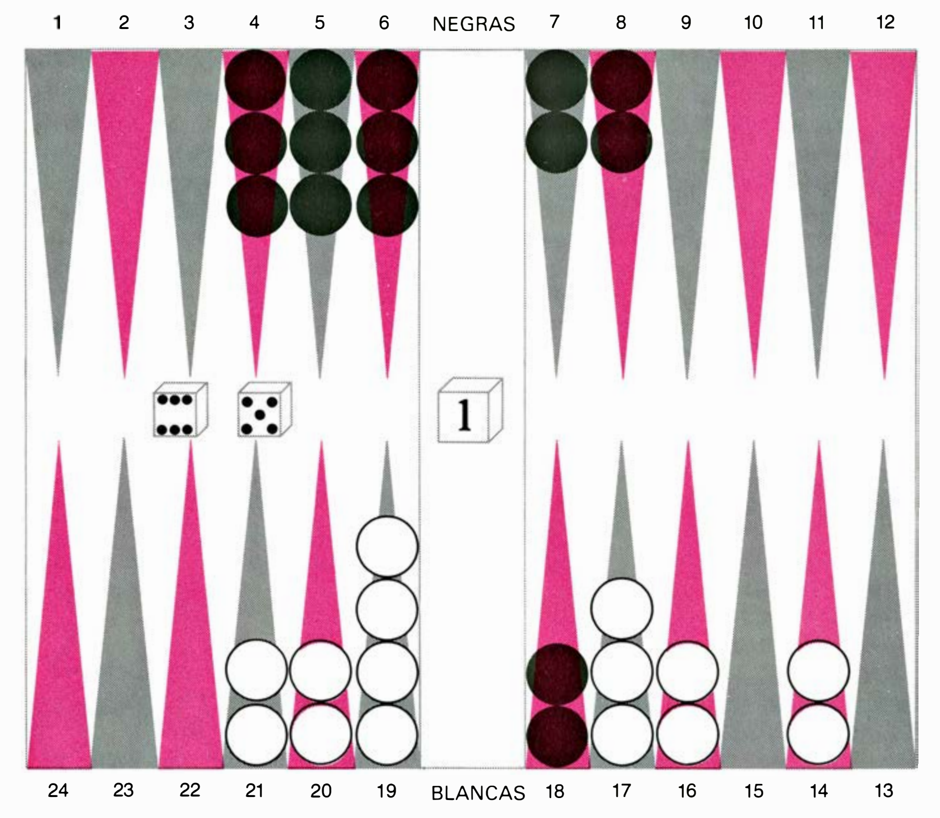
La construcción de funciones de evaluación eficaces exige ensayar en programas reales diversos enfoques del problema. La función de evaluación de mi primer programa se basaba en un polinomio de primer grado, donde cada término representaba una determinada faceta del momento de juego, y el coeficiente del término medía la importancia relativa de tal característica. La suma del polinomio expresaba el valor concedido a la posición. Este enfoque presentaba una limitación, a saber, que siendo constantes los coeficientes, éstos solamente podían expresar la importancia que en promedio revestía tal característica. Esta limitación solía impedir al programa formular apreciaciones verdaderamente dignas de expertos.



LOS FINALES DE PEONES son situaciones típicas donde los programas de ajedrez se desenvuelven mal. La que vemos da buena muestra del llamado "efecto de horizonte": el programa cree haber resuelto un problema, cuando en realidad tan sólo ha conseguido posponerlo. El programa, que juega con blancas, fue diseñado para que eligiera una jugada tras examinar todas las variantes hasta profundidad de tres pasos (dos de blancas y uno de las negras). El programa encuentra, entre otras, la continuación 1. P8T (D), P8A (D); 2. D7C +, y estima atinadamente que de esta forma lleva ventaja de dos peones, más la iniciativa. Empero, el programa descubre también que jugando 1. P6D +, RxP; 2. P8T (D) llevaría una reina de ventaja, y por ello, el programa opta por esta segunda continuación, sin percatarse de que en la siguiente jugada las negras podrían coronar un peón. La causa está en que se le ha ordenado explorar tan sólo hasta profundidad de tres turnos. Una jugada más tarde el programa se apercibiría de que la jugada prevista, P8T (D), no le otorga la ventaja esperada, a causa de P8A (D) en las negras; por ello vuelve a servirse de un remedio artificial: 1. P5R+, RxP; 2. P8T (D). Para jugadores humanos todo este planteamiento es absurdo, porque enfocan el juego bajo distinta perspectiva: no se interesan por jugadas, sino por acontecimientos. La coronación de un peón negro es acontecimiento importante, que el programa supo detectar. Mientras en alguna rama del árbol de juego quede algún acontecimiento importante por explorar, los jugadores humanos, a diferencia de la máquina, persisten en el análisis.

Mi siguiente intento consistió en dividir las posiciones del juego en varias categorías, cada una dotada de distintas funciones de evaluación. Este nuevo enfoque reforzó la potencia de juego del programa, pero también tenía un inconveniente. Los valores relativos de dos posiciones pertenecientes a distinta clase eran a veces imprecisos, porque las funciones de evaluación producían valores significativamente diferentes para situaciones de juego cercanas a la divisoria entre clases, mientras que, por el contrario, tales valores hubieran debido ser muy parecidos.

Por ejemplo, supongamos que el programa deba optar entre una jugada pensada para bloquear al contrario (donde los elementos a tomar en cuenta son completar puntas, proteger peones aislados, atacar peones vulnerables del contrario, etcétera) y una jugada pensada para la carrera (donde los ejércitos contrarios ya se han rebasado uno al otro, y la única consideración es cuál de ambos lleva delantera en la carrera hacia la salida). Por otra parte, en el juego de bloqueo cada bando debe estimar cuál sería el resultado de la carrera que se produciría si ambos ejércitos se destrabasen súbitamente. Dado que en la fase de carrera la partida se desarrolla de forma totalmente distinta de la de bloqueo, el programa tropieza con dificultades para comparar ambas. Lo abrupto de la frontera entre ambas clases de juego, que ocasionalmente produjo groseros errores de enjuiciamiento, parecía casi imposible de remediar.



SITUACION DE CHAQUETE, típica de la que las primeras versiones del programa BKG no sabían analizar correctamente. Las negras acaban de sacar un seis y un cinco, que pueden utilizarse en dos formas esencialmente distintas: mover las piezas de la punta 18 (18-12; 18-13), o bien, mover otras dos piezas. La primera estrategia (que provocaría una carrera hacia la salida) destraba las piezas de ambos bandos, y lo que resta de partida será carrera pura y simple. La segunda, conducente a juego de bloqueo, deja trabados a los ejércitos de piezas contrarias; en esta situación tiene gran importancia la ocupación de puntas, la protección de piezas en precario, el ataque a piezas contrarias, etcétera. El programa podría haber comparado eficazmente dos alternativas de un juego de carrera, o dos jugadas de bloqueo. Mas como los juegos de carrera y de bloqueo son muy diferentes, el programa tuvo dificultad para decidir qué formación le es más conveniente. El éxito del BKG 9.8 obedece a que opera por evaluación posicional.

Un nuevo intento de obtener funciones de evaluación adecuadas resultó mucho más fecundo. Para superar el problema de las fronteras abruptas se optó por no dividir las posiciones en clases. Lo que se hizo, en cambio, fue deformar el espacio de evaluación, constituido por todas las posiciones, de forma que ciertas partes del espacio dieran más importancia a unas características que a otras. Se hizo que las transiciones de importancia de unas partes a otras fuese progresiva. Además, la transición dependía de otras características presentes, lo que significa que la importancia de una determinada característica está dada por una función no lineal. Las características que controlan las transiciones se llaman coeficientes de aplicación; son variables que cambian lentamente y reemplazan a los coeficientes, normalmente constantes, del polinomio lineal de evaluación. Como los coeficientes de aplicación varían lenta y progresivamente, reflejan gran cantidad de contexto, pero evitan al mismo tiempo crear lindes abruptos entre distintos contextos. Este nuevo enfoque fue bautizado SNAC, siglas de progresividad, no linealidad y coeficientes de aplicación, en inglés.

El enfoque SNAC permitió por vez primera que el programa BKG (fuga de vocales de Backgammon, nombre inglés del chaquete) hiciera buen uso del inmenso caudal de conocimientos que por entonces le habíamos infundido ya. Las anteriores versiones del programa habían padecido muchas limitaciones por aplicar de forma demasiado rígida tales conocimientos. Cada función SNAC es, en efecto, una microestrategia para acentuar o eludir ciertos rasgos de las posiciones del juego. El coeficiente indica la medida en que es aplicable la estrategia. En consecuencia, el programa BKG puede examinar simultáneamente un número cualquiera de las aproximadamente 30 estrategias para las que tiene dispuestas funciones SNAC, guardando constancia al mismo tiempo de la importancia de cada estrategia en esa posición concreta. En un momento dado del juego, ciertas estrategias pueden quedar relegadas por completo, otras potenciadas a su máximo grado, y otras, recibir solamente atención moderada. Este enfoque permite desplazar gradualmente el centro de gravedad al atribuir pesos a las estrategias, sin modificaciones repentinas y aberrantes del rendimiento de juego al irse presentando nuevas exigencias.

La estructura y contenido del programa fueron evaluados, en condiciones controladas, de tres maneras: proponiéndole problemas tomados de manuales de iniciación; contrastándolo en enfrentamientos contra otros programas; y enfrentándolo a jugadores humanos. Los resultados del programa BKG, tanto al resolver problemas de manual como en partidas auténticas, reflejaron que la introducción de funciones SNAC supuso una notable mejora, pese que el programa apenas si había reforzado sus conocimientos del juego. En un enfrentamiento, el programa BKG le zurró al mejor microprocesador comercial disponible en el 78 por ciento de las veces. Sin las SNAC, el BKG solamente consiguió vencer el 56 por ciento. La primera vez que Magriel jugó contra el programa quedó sorprendido de lo bien que se defendía. Posteriormente, Magriel nos ayudaría a refinar los algoritmos de gestión del cubo de duplicar y a evaluar el rendimiento global del programa.

La prueba de fuego de la potencia real del BKG tuvo lugar en mayo de

	ANTES DE LOS SNAC	CON SNAC
PROBLEMAS DE MANUAL	45%	66%
CONTRA EL MEJOR MICROPROCESADOR COMERCIAL	56%	78%
CONTRA ANTERIORES VERSIONES DEL PROGRAMA BKG	37.9%	62.1%
CONTRA JUGADORES HUMANOS	0-2	3-1

ESTA TABLA DE COMPARACION DE RESULTADOS muestra cómo el programa obtuvo resultados muy superiores cuando la función polinómica utilizada para estimar la fuerza de una situación fue dotada de funciones y coeficientes de aplicación no lineales y de variación suave (denominados SNAC). En el polinomio cada sumando refleja un rasgo de la situación de la partida, y su coeficiente, de carácter no lineal, sirve para ponderar la importancia de tal aspecto. Por no ser los coeficientes constantes, sino variables, la importancia de cada rasgo particular depende de otras características de la situación.

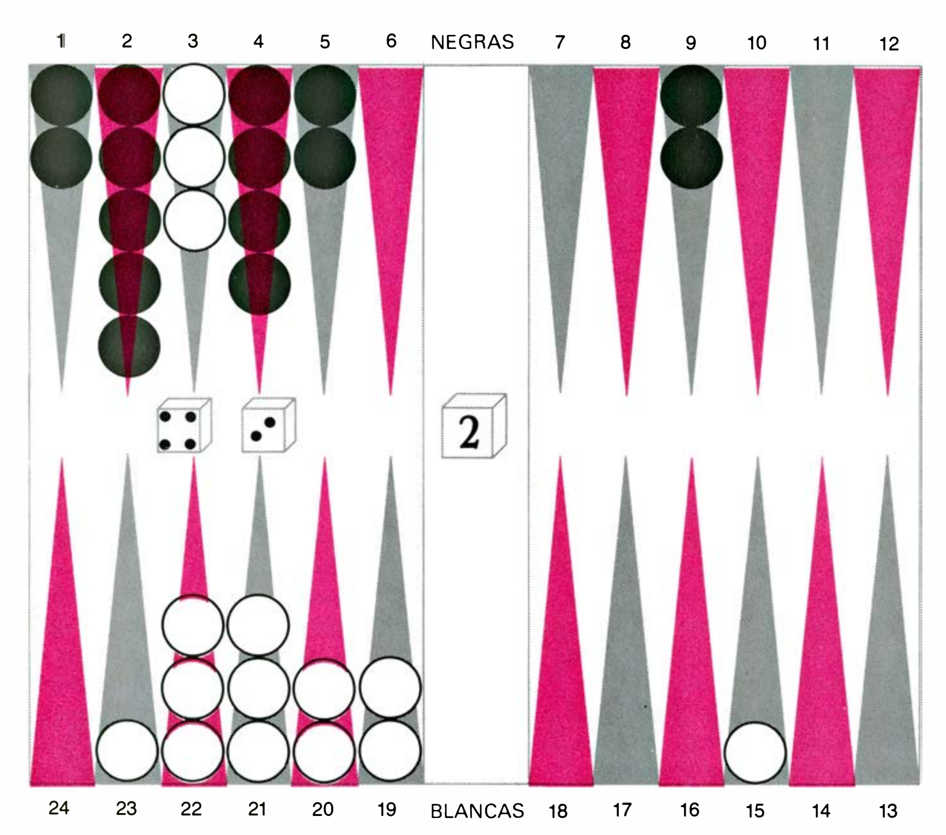
1979, cuando el BKG 9.7 (los números denotan las sucesivas versiones del programa) intervino en un pequeño campeonato entre particulares, jugadores de nivel intermedio, en California, ganando los dos primeros encuentros antes de caer frente al que sería campeón definitivo del torneo. Sin SNAC, el programa no se había defendido bien en campeonatos. En la primavera de 1978, el programa quedó eliminado en un campeonato interno de la Carnegie-Mellon University, perdiendo sus primeros encuentros contra jugadores de poca talla.

Sólo faltaba un mes para la semana decisiva, en Monte Carlo, cuando el programa tendría que vérselas con el nuevo campeón del mundo. Antes de que los contrincantes pudieran enfrentarse, los mejores jugadores humanos habrían de vérselas entre sí, con los ojos puestos en el campeonato mundial, que finalmente ganaría Luigi Villa.

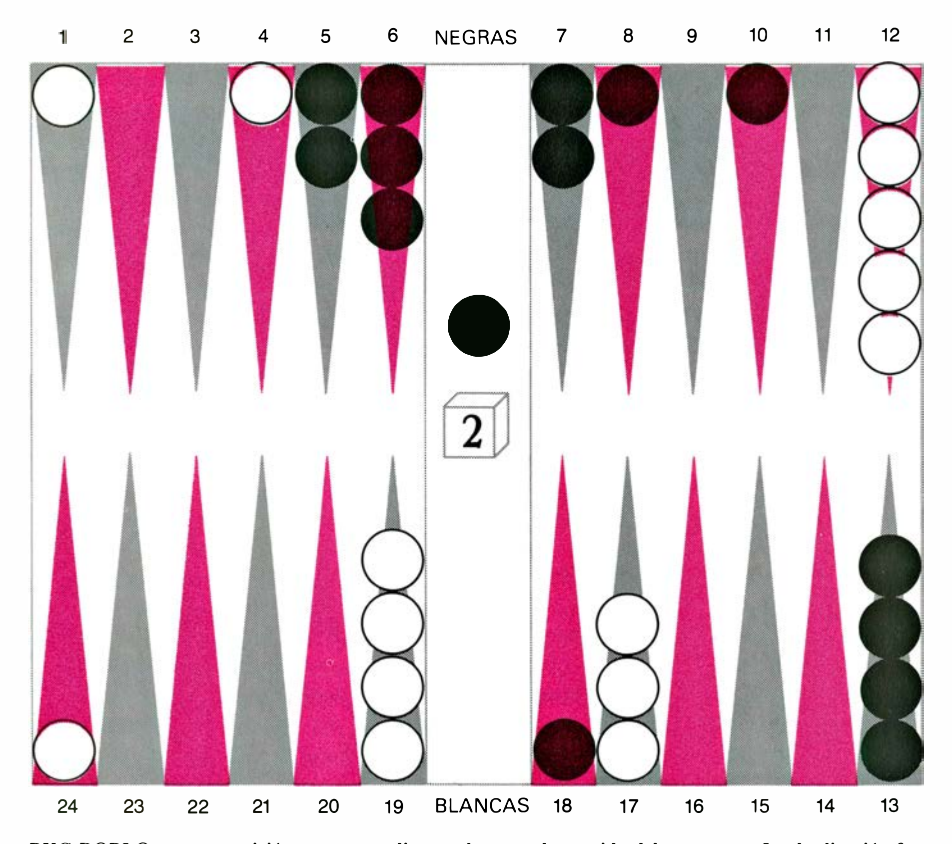
El programa BKG 9.8, que operaba en un ordenador PDP-10 de la Carnegie-Mellon University, fue enlazado vía satélite con un robot móvil situado en Monte Carlo. No se esperaba gran cosa del robot programable, que los organizadores del torneo apodaron "Gammonoid". Pese a que los organizadores habían hecho de nuestro Gammonoid el símbolo del torneo, imprimiendo fotografías suyas en los folletos informativos y grabando pequeñas imágenes del robot en los trofeos, los participantes sabían que los microprocesadores existentes no les presentarían dificultades serias. ¿Por qué habría de ser diferente el robot?

Esta impresión quedó reforzada en las ceremonias de apertura, en el Palacio de Deportes de Verano de Mónaco. En cierto momento las luces de la sala se atenuaron, la orquesta atacó el tema de *La guerra de las galaxias,* y un proyector dibujó un círculo de luz en una abertura del telón del proscenio, por donde se esperaba saliese al escenario el Gammonoide. Para desilusión mía el robot se enredó en la cortina, y su presentación se retrasó cinco minutos.

El duelo entre BKG 9.8 y el campeón tuvo lugar a las 11 de la noche del último día del torneo. Cada bando aportó 2500 dólares a fin de que el encuentro tuviera mordiente. Para convencer a los espectadores que no había trampas en el lanzamiento de los dados para el programa, el director del campeonato nombró un ayudante, que los lanzaría para el robot. El encuentro, que se celebró en el Palacio de Depor-



BKG SACO UN CUATRO Y UN DOS encontrándose en esta situación, de la primera partida contra Villa. El programa, que juega con negras, tenía ventaja, pero se vio obligado a dejar un peón en precario. Optó por jugar 9-5 y 9-7, dejando un peón precario en la punta 7, que podría ser alcanzado por 13 posibles lanzamientos de los dados. Una jugada en apariencia más segura sería 5-1 y 4-2, dejando el peón precario en la punta 5, que sólo podría ser alcanzado de 11 maneras. Pero esta jugada es inferior a la realizada, porque deja dos piezas en la punta 9, que podrían quedar desprotegidas cuando hubieran de moverse. BKG dispone de una función que cuenta el número de formas en que un peón es atacable.



BKG DOBLO en esta posición, correspondiente a la segunda partida del encuentro. La duplicación fue muy atinada, pues Villa optó por rehusarla. El problema de duplicar se reduce a decidir si la posición propia es demasiado débil, demasiado fuerte o la justa. Si la posición es demasiado débil, el oponente aceptará la duplicación. Si es demasiado fuerte, la rechazará, perdiendo tan sólo la puesta inicial, y no el doble, como le sucedería de no haberse ofertado la duplicación y haber sido derrotado por chaquete.

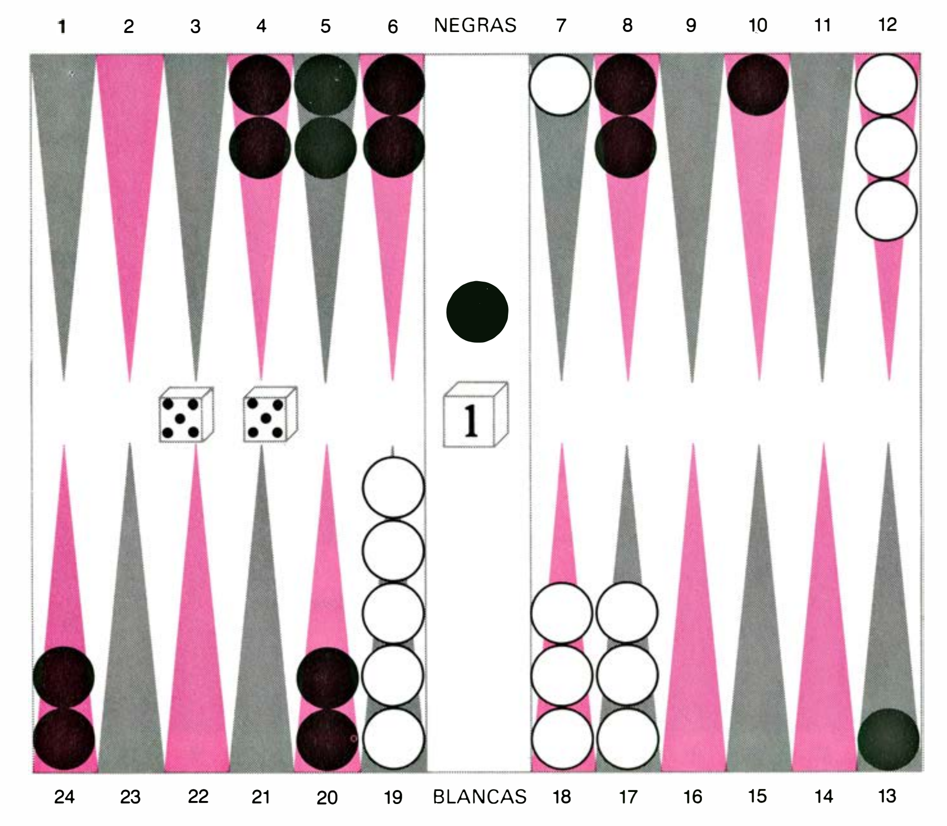
tes de Invierno, fue retransmitido por circuito cerrado de televisión a una sala capaz para 200 personas. Pese a que Magriel estaba preparado para ir comentando las jugadas la concurrencia fue desalentadoramente reducida.

En la primera partida, el BKG 9.8 consiguió salir con buen pie. Duplicó la puesta pronto, cuando Villa no tenía más remedio que aceptar, y hábilmente fue engrosando su ventaja, hasta alcanzar una posición de casi seguro vencedor. La ilustración superior de la página precedente muestra una situación crucial, a punto de terminar la partida. Al lanzar los dados, el programa sacó un cuatro y un dos, y estaba obligado a dejar un peón en precario. ¿Pero cuál? Podía mover una pieza desde la punta 5 hasta la 1, y una pieza desde la 4 hasta la 2, dejando un peón vulnerable en la punta número 5, que podía ser atacada de 11 maneras distintas (mediante cualquier tirada de los dados que contuviera un dos). El programa optó por mover dos piezas, desde la punta 9 a las números 5 y 7, dejando un peón vulnerable en la punta 7, alcanzable de 13 maneras (cuando los dados mostrasen al menos un cuatro, o bien, un tres y un uno).

El programa BKG dispone de una función que cuenta el número de formas en que un peón aislado es atacable, y, por ello, es perfectamente capaz de calcular el riesgo de sufrir una baja que estaba corriendo. El programa dispone también de funciones SNAC que miden el riesgo futuro de sufrir bajas, de acuerdo con las diversas clases de formaciones y de la probabilidad de que dichas formaciones se descompongan inmediatamente. Las dos jugadas entre las que el programa tuvo que decidir difieren en el balance entre riesgos inmediatos y riesgos futuros; pero las funciones SNAC están bien calibradas para abordar situaciones de este tipo, y por ello el programa no tiene dificultad para decidir qué jugada es preferible. El programa se da cuenta de que ha elegido la jugada de mayor riesgo inmediato. Empero, se da cuenta también de que, aún jugando a lo más seguro, pronto hubiera tenido que mover las piezas situadas en la punta 9. Probablemente hubiera tenido entonces que dejar un peón vulnerable, pues la próxima punta amiga está alejada de la número 9. En consecuencia, el programa prefiere correr un riesgo inmediato algo mayor, pero liquidar al mismo tiempo el problema a largo plazo. Era sin duda la jugada correcta, la que cualquier jugador experto efectuaría sin pensar, pero no era el tipo de cosas que se considera factible mediante programas de ordenador; como, por ejemplo, adoptar un curso en apariencia imprudente porque a largo plazo puede dar mejores resultados. Tras esta jugada de BKG, Villa no tuvo la fortuna de alcanzar el peón en precario,y el programa consiguió llevar la partida a buen puerto, venciendo por 2-0.

La segunda partida era en realidad el eje del encuentro. El programa consiguió hacer otra buena salida, y para conservar alguna probabilidad de victoria, Villa atacó una de las piezas de BKG de su propia mesa interior. Se produjo así la situación que vemos en la ilustración inferior de la página precedente. Es el turno de juego del programa. Ahora BKG ofreció duplicar la puesta, jugada extraordinariamente imaginativa, que Villa rehusó. El programa había apreciado que disponía ya de posición ligeramente superior al contrario, y que un golpe de suerte al lanzar los dados (cualquier tirada que le diese un uno, concretamente, cualquier tirada que le permitiese atacar la pieza aislada en la punta 4) podría fácilmente tornar el juego muy a su favor. Y entonces sería demasiado tarde para doblar, porque Villa rehusaría, impidiendo al BKG la posibilidad de una victoria doble. Los especialistas aplaudieron la duplicación, si bien consideraron que Villa debió aceptarla. En realidad, Villa declinó porque confiaba en llevar la voz cantante en partidas sucesivas. La puntuación era ya de 3-0, y la sala de televisión comenzó a animarse.

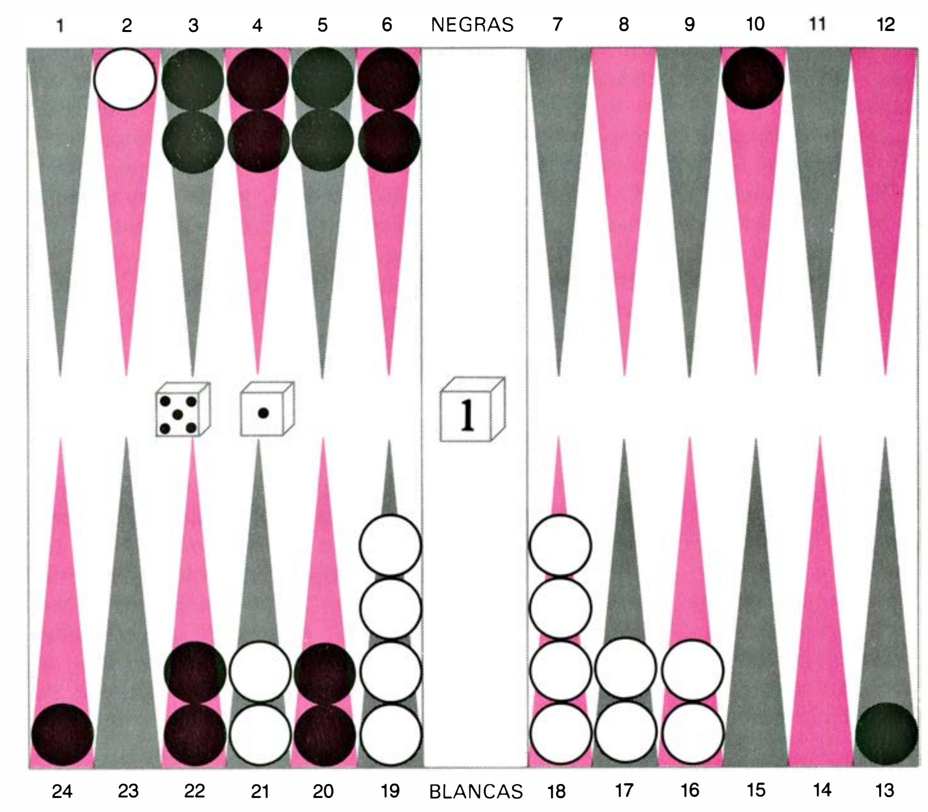
Villa consiguió tomar delantera en la tercera partida. Construyó una sólida posición y dobló en un momento oportuno para que BKG declinase. Empero, el programa aceptó el envite. Villa tenía ahora oportunidad de ganar un juego doblado, pero una serie de lanzamientos desafortunados le obligaron a dejar en precario un peón cuando ya estaba retirando piezas del tablero. El programa no pudo atacar inmediatamente ese peón, pero dos jugadas más tarde Villa se vio obligado a dejar en precario otro peón. Esta vez, BKG recibió los favores de la suerte, alcanzó una de las piezas de Villa y consiguió llegar a vencerle, redoblando cuando su posición era ya desesperada. La puntuación era entonces de 5-0, y la sala de televisión estaba repleta de espectadores que rezongaban y animaban conforme las alternativas del juego favorecían a uno u otro combatiente. Era como en las semifinales y la final del campeonato del mundo, con la sala atiborrada de aficionados que discutían qué jugadas serían preferibles, y qué bando llevaba la ventaja.



BKG SACO DOS CINCOS en la quinta y última partida del encuentro. El programa jugó 8-3, 8-3, y barra-15, dejando tres peones en precario. Si alguno de ellos fuese alcanzado, BKG dispondría de más tiempo para reforzar su juego trasero, a saber: dominar las puntas 24 y 20 de la mesa interior de Villa.

En este punto yo estaba encantado por el cariz que tomaban los acontecimientos. En mis más íntimos y felices sueños yo había esperado perder el torneo, aunque consiguiendo 5 o 6 puntos. Empero, henos aquí con una ventaja de 5-0 y excelentes probabilidades de vencer el encuentro. Yo había temido que un experto, jugando de forma imprevisible, consiguiera hacer caer al programa en celadas donde, desconcertado, tal vez cometería disparates. Evidentemente, trabajamos mucho para protegerlo de tal posibilidad. No obstante, la temía en mi fuero interno, aunque no llegara a presentarse.

En la cuarta partida, Villa sacó ventaja nada más empezar, proponiendo una duplicación casi inmediatamente, confiando en que si conseguía vencer podría reponerse de sus pérdidas. Sin embargo, el programa decidió ir sobre seguro y conservar lo más posible su posición dominante en el encuentro, y declinó la duplicación. Fue decisión debatida si el programa debería haber aceptado o no. La puntuación era ahora de 5-1 y la muchedumbre que llenaba la sala de televisión confiaba en que Villa se rehiciera.



BKG OBTUVO UN CINCO Y UN UNO en este momento de la última partida. El programa realizó la sensacional jugada 13-8 y 3-2. De ser alcanzada alguna de sus piezas vulnerables, el programa dispondría de más tiempo para reforzar su juego trasero. Por otra parte, de no serlo ninguna, podría ir llenando puntas de su cuartel de llegada, dificultando a la pieza de Villa reunirse con las suyas y escapar.

La quinta partida fue con mucho la más interesante y trascendente del encuentro. El programa jugó con enorme agresividad en las fases iniciales, y en consecuencia, dos de sus piezas fueron alcanzadas y tuvieron que volver a empezar. El programa sacó el máximo partido de esta pérdida de tiempo, planteando un juego trasero, defensivo, extraordinariamente sólido. En él, las puntas defensivas se mantienen en el cuartel de llegada del contrario, lo que implicaba que la victoria de Villa, si llegaba a producirse, estaba todavía lejana. La idea central del juego trasero consiste en tentar al contrario a avanzar rápidamente, con la esperanza de poder atacarle alguna pieza y enviarla al punto de partida. Para que esta estrategia tenga éxito es necesario disponer de una sólida posición defensiva, con las piezas propias no demasiado adelantadas; de lo contrario no será posible retener las piezas retrasadas del contrario.

Ambos bandos jugaron bien durante varias tiradas; entonces, al lanzar el BKG obtuvo un desdichado cinco doble, cuando la situación estaba como vemos en la ilustración de la página anterior. En consecuencia, el programa tuvo que adelantar sus piezas más rápidamente de lo que hubiera querido. Demostró su excelente dominio de la posición, moviendo dos piezas desde la punta 8 hasta la 3, más una pieza desde la barra hasta la punta 15, dejando así tres peones vulnerables. Evidentemente, al programa poco podía importarle que algunos o todos ellos fuesen alcanzados; ello hubiera retrasado el avance de las piezas propias y lo habría situado en mejor posición para atacar alguna pieza de Villa. Además, las piezas en precario estaban situadas de forma perfecta para obstruir el avance de Villa, quien no se podía permitir sufrir bajas, porque el programa disponía ya de una buena formación defensiva.

Algunas tiradas más tarde, BKG tuvo que jugar un cinco y un uno cuando se encontraba en la posición que vemos en la figura de esta página. Saltaba a la vista que debería trasladarse una pieza de la punta 13 a la 7, con la esperanza de que esta pieza fuese alcanzada cuando la última pieza de Villa, atrapada en la punta número 2, intentase escapar. De resultar alcanzada la pieza, el programa dispondría de más tiempo para preparar su defensa. Sin embargo, el programa despreció tal jugada, realizando en cambio otra sensacional, pasando un peón de la punta 13 a la 8, y otro de la número 3 a la número 2.

A pesar de que así quedaba destruida parte de la formación defensiva del BKG, la jugada era una agresiva espada de doble filo, que los observadores expertos juzgaron atinada. La explicación es como sigue: si Villa atacase uno o ambos peones del cuartel de llegada del BKG, éste ganaría tiempo para mejorar la eficacia de su juego trasero. Por otra parte, si dichos peones no fueran alcanzados, BKG tendría posibilidad de plantear una posición defensiva todavía más sólida, dificultando todavía más la huida de la pieza retrasada de Villa de las piezas del programa que la retienen confinada. En esencia, la sobresaliente jugada de BKG creaba no una, sino dos vías hacia la victoria. No obstante, la jugada era arriesgada, y Villa, con mucho que ganar y nada que perder, ofreció duplicar. BKG aceptó.

Villa alcanzó inmediatamente una pieza al adelantar la suya con la intención de agrupar sus fuerzas en la meta. El programa lanzó entonces un par de cincos, que le obligaban a moverse más rápidamente de lo deseado. Pocas jugadas más tarde se presentaba otra situación crítica. BKG tenía que jugar dos doses cuando se encontraba como vemos en la ilustración de la página siguiente. Aquí un jugador no muy experto hubiese estado tentado de mantener su juego trasero, moviendo una pieza de la punta 22 a la 20 y dos piezas de la punta 6 hasta las puntas 4 y 2. No obstante, BKG se dio cuenta,

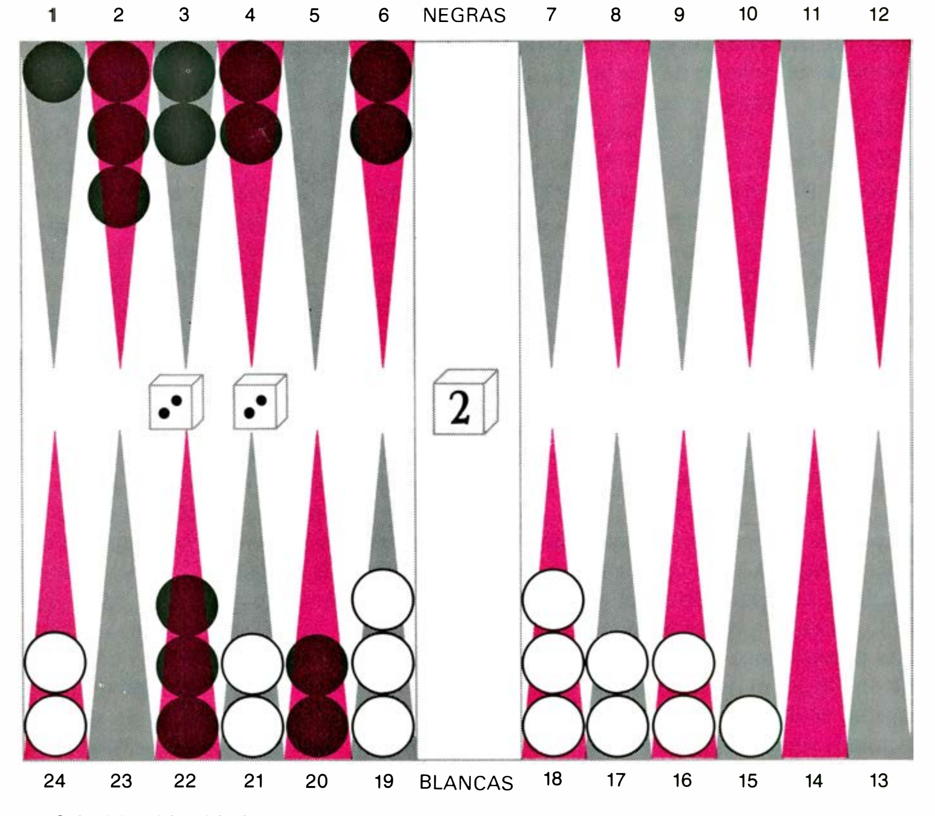
como harían casi todos los especialistas, de que el juego trasero tiene poco valor sin una sólida situación en el cuartel de llegada. Más todavía, de jugarlo a ultranza se corre el riesgo de sufrir chaquete en una partida doble.

Por consiguiente, BKG llevó una pieza de la punta 3 a la número 1, y tres piezas desde la 22 a la 20. Con estos movimientos se abandonaba la estrategia de juego trasero y se comenzaba la carrera hacia el cuartel de llegada, para limitar las pérdidas a las de un juego doblado. Pronto el BKG recibiría recompensa por su buen jugar. En los 12 turnos siguientes de cada bando, el encargado de lanzar los dados para el programa sacó a Villa 31 puntos de ventaja (la desviación típica al lanzar los dados una vez es 4.3 puntos), consiguiendo así una posición que ya tenía alrededor del 20 por ciento de probabilidad de victoria. Y en ese momento salieron del cubilete dos seises, que permitieron al programa ganar la partida y el encuentro.

A duras penas podía yo creer este resultado final, y eso que al programa no le habían regalado la victoria. No hubo en su juego ningún error serio, aunque desde luego tuvo suerte al ganar las partidas tercera y final. Hubo fotografías y entrevistas; los expertos allí reunidos me felicitaron con calor. Sólo una cosa echaba a perder la escena. Villa, quien sólo un día antes había alcanzado el cenit de su carrera como jugador de chaquete, quedando titular del campeonato mundial, estaba desconsolado. Le expliqué que lamentaba lo sucedido, y que ambos sabíamos que en realidad era él quien mejor jugaba.

Algunas semanas después, ya en la Carnegie-Mellon, analicé las partidas con cierto detalle. No cabía duda de que BKG había jugado bien, pero en el fondo, Villa jugó aún mejor. Villa realizó prácticamente siempre las jugadas técnicamente correctas, mientras que el programa dejó de efectuar la jugada óptima en ocho de las 73 ocasiones en que su movimiento no fue forzoso. Empero, tan sólo uno de estos errores llegó a causarle auténticas dificultades. Un experto no hubiera cometido casi ninguno de los errores del programa, si bien éstos solamente pudieron ser explotados en un reducido tanto por ciento de veces.

Donde BKG brilló con fuerza fue en el ejercicio de la imaginación. Quien no supiera el color de las piezas del programa hubiera podido pensar que era el jugador humano quien ejecutaba jugadas espectaculares, y la máquina, las trilladas y rutinarias. En el encuentro, los dados tuvieron la última palabra y se pronunciaron por BKG 9.8.



BKG SACO DOS DOSES en esta posición de la partida definitiva. Al efectuar las jugadas 3-1, 22-20, 22-20, y 22-20, el programa actuó con gran tino, abandonando el juego trasero y emprendiendo la carrera hacia el cuartel de llegada, con la intención de evitar sufrir chaquete. Pese a que en la carrera el programa llevaba las de perder, tuvo mucha suerte con los dados y terminó por ganar esa partida.

El programa BKG 9.8 será ahora desarrollado en dos aspectos. Ante todo, me propongo darle más conocimiento y experiencia. Y segundo, proyecto convertirlo en un programa de aprendizaje, capaz en última instancia de percibir sus propias flaquezas y corregirlas. La primera tarea llevará tiempo, pero no parece que vaya a crear dificultades imprevistas. El programa BKG 9.8 parece conocer ya casi todo lo fundamental en chaquete. Aunque la precisión de su conocimiento es bastante elevada, hay aspectos donde los valores de los coeficientes de aplicación podrían mejorarse algo, y otros donde mayores conocimientos podrían ser beneficiosos. Mis colaboradores y yo estamos trabajando actualmente en estas cuestiones.

Muchas cuestiones del chaquete, como por ejemplo el método óptimo de ir retirando piezas del tablero, pueden ser totalmente resueltas mediante análisis numérico, con ayuda de ordenadores. Uno de mis discípulos ha puesto a punto una base de datos completa para todas las posiciones con 15 piezas o menos en el cuartel de llegada. La base de datos informa del número esperado de lanzamientos (aproximado hasta la quinta cifra decimal) necesarios para conseguir retirar del tablero todas las piezas. A partir de esta información es posible determinar la jugada óptima correspondiente a cualquier puntuación de los dados y cualquier disposición de las piezas. No obstante, la base de datos contiene 54.000 ítems, y sería un despilfarro de recursos del ordenador hacerla formar parte del programa. De todos modos, parece que será posible preparar unas cuantas reglas que resuman todo cuanto sabemos sobre retirada de piezas. Nos encontramos ahora en el proceso de poner a punto un programa capaz de generar sus propias reglas, ensayarlas y modificarlas hasta que finalmente consiga una reducida lista de ellas que recojan toda la información de la base de datos. De lograrlo, tales reglas pasarían a formar parte del programa BKG.

Otro problema resoluble se refiere al momento en que es oportuna la duplicación. En último análisis, el problema se reduce a decidir si la propia posición es demasiado buena, si no es lo suficientemente buena, o si es lo justamente buena para ofrecer duplicación. La cuestión crucial consiste en estimar cuánto y en qué sentido habrá evolucionado la situación para cuando se tenga la próxima oportunidad de envidar doble. Puesto que hay 21 resultados posibles del lanzamiento de los

dados para cada bando antes de que se presente la siguiente oportunidad de duplicar, es preciso examinar 441 situaciones posibles, suponiendo, claro está, que sea factible evaluarlas todas ellas. Por lo general, tales situaciones pueden organizarse en grupos; se evalúa entonces cada uno de éstos. Si resulta que la posición propia sería demasiado fuerte para que conviniera ofrecer duplicación en la siguiente oportunidad en la mayoría de las 441 situaciones posibles, entonces la duplicación debe envidarse inmediatamente. Por otra parte, si las probabilidades fueran que el contrario estaría igualmente bien dispuesto a aceptar el envite tanto ahora como en la próxima jugada, entonces se debe esperar, porque puede suceder que a él le sonría la suerte y saque con los dados buenas puntuaciones, mientras que las propias pueden ser malas. Los jugadores humanos expertos enfocan el problema de la duplicación desde ángulo muy próximo a como lo hace el BKG. De todas formas, si el programa recibiera entradas precisas y fiables para sus ecuaciones de duplicación, en esta fase del juego podría desbancar a los jugadores humanos.

A veces el chaquete presenta situaciones para las que no disponemos de modelos bien perfilados. Al hablar de tales situaciones suele decirse que requieren buen juicio, sentido común. El programa se defiende francamente bien en este dominio a causa del éxito de las funciones SNAC. No obstante, esperamos perfeccionarlo en varios aspectos. Me propongo primero generar gran número de situaciones parecidas, que requieran juzgar con tino, y después, retocar, afinar y aumentar las funciones afectadas hasta que generen respuestas correctas en todos los casos. Esta tarea podría resultar penosa, y sería preferible realizarla automáticamente, pero no sé todavía cómo poder conseguirlo.

Parece que las funciones SNAC serán el medio idóneo para recoger las características que los humanos calificamos de buen juicio. Tales funciones permiten responder a leves variaciones de los estímulos con pequeñas variaciones de conducta, y esto es precisamente lo que separa al buen juicio de la deducción lógica. Soy de la opinión de que las funciones SNAC podrían aprovecharse ventajosamente en aplicaciones computerizadas, que podrían ir desde cuestiones de estética hasta la elaboración de decisiones en la vida comercial ordinaria. Evidentemente, para concretarlas de forma razonable será necesario detectar antes las variables importantes y las relaciones que las ligan, y ajustar a ellas las funciones necesarias.

USUARIO:	"¿En qué etapa del juego nos encontramos?"
PROGRAMA:	"El juego está en fase de salida del tablero. Lleva Vd. la delantera".
USUARIO:	"¿En qué consiste mi ventaja?"
PROGRAMA:	"Lleva Vd. delantera en la carrera de salida".
USUARIO:	"¿Qué debo tratar de hacer?"
PROGRAMA:	"Debería Vd. eliminar sus barreras, y tratar de llevar sus piezas al cuartel de llegada sin correr riesgos".
USUARIO:	"¿Qué quiere decir «barrera»?"
PROGRAMA:	"El grado en que están encerradas las piezas de su contrario".
USUARIO:	"¿De qué depende esto?"
PROGRAMA:	"Del número de piezas situadas frente a un conjunto de piezas que Vd. domina, y de la dificultad de rebasar dichas puntas".
USUARIO:	"¿Qué fuerza tiene mi barrera?"
PROGRAMA:	"Es moderadamente fuerte".

FUTURO DIALOGO entre un usuario humano y un programa BKG. El jugador humano seleccionará sus preguntas de una lista, y el jugador programa ofrecerá respuestas examinando su propia base de conocimientos y el conjunto de posibilidades implícitas en la correspondiente situación de juego.

Es muy posible que durante los próximos cinco años veamos enormes avances en métodos de juego por ordenador. La era actual, donde existen ya potentes programas de juego, empezó en julio de 1976, cuando un programa llamado Chess 4.5, diseñado y puesto a punto por David J. Slate y Lawrence R. Atkin, de la Northwestern University, venció en un campeonato entre jugadores de Categoría *B* celebrado en California. (En la escala oficial de clasificación norteamericana, la Clase *E* va desde 0 hasta 1199 puntos, la clase *D*, de 1200 a 1399, la Clase *C*, de 1400 a 1599, la Clase *B*, de 1600 a 1799, la Clase *A*, de 1800 a 1999, la categoría de expertos empieza en 2000, hasta 2199, y la categoría máxima, maestro, va de 2200 en adelante. La mediana de las puntuaciones de los participantes en torneos en los Estados Unidos es 1362 puntos.) Desde aquella fecha, el programa Chess 4.5 ha ganado el torneo "open" de Minnesota, y en partidas "relámpago" (donde ningún bando puede consumir más de 5 minutos en toda la partida) ha dado zurras a expertos y maestros.

Otro programa muy potente para jugar al ajedrez es el llamado Belle, preparado por Ken Thompson y Joe Condon, de los Laboratorios Bell, y que utiliza circuitos especiales diseñados específicamente para analizar situaciones de ajedrez. Pronto será reforzado, haciéndolo capaz de examinar muchos millones de nudos en el margen de tres minutos por jugada, lo que produce pasmo a la imaginación. Cuando se combine el saber del Chess 4.5 con la circuitería reforzada del Belle, el programa, confío yo, será capaz de jugar al nivel de gran maestro. Y quizá las funciones SNAC ayuden al programa a desarrollar más fina sensibilidad hacia los elementos importantes de cada posición del juego.

Todavía más brillante parece el futuro del chaquete por ordenador. Mi programa juega en categoría *A*, nivel avanzado intermedio. Cuando se le introduzcan las mejoras que he reseñado espero que será capaz de jugar en todas las fases del juego al nivel de experto, e incluso sabrá hacer ciertas cosas mejor que cualquiera de sus contrincantes humanos desprovistos de ayudas mecánicas o memorísticas, si se les limita el tiempo. En este aspecto, incluso a los mejores jugadores ha de resultarles difícil vencer el programa. Confío en llegar a transformar el programa BKG en un programa autodidacta, capaz de perfeccionarse estudiando sus propias partidas.

Muchas personas consideran que juegos como el chaquete o el ajedrez son intrínsecamente frívolos. Pero los resultados alcanzados en juegos mediante ordenador son sin duda comparables con cualesquiera otros conseguidos en el campo de la inteligencia artificial. Los trabajos sobre ajedrez por ordenador han producido avances fundamentales en técnicas de exploración de árboles muy ramificados, y para fragmentar tareas de este tipo, analizándolas por computación en paralelo. Y el trabajo sobre chaquete ha permitido dar con mejores métodos para organizar conocimientos y enjuiciar situaciones.

Ciencia y sociedad

Amniocentesis genética

Se llama amniocentesis a la inserción de una aguja en el saco membranoso que rodea al feto en la matriz de una mujer embarazada y la extracción de cierta cantidad de líquido en el que está inmerso el feto. En 1955 Povl Riis y yo, trabajando en la Universidad de Copenhague, demostramos que el sexo de un feto podía determinarse tiñendo los núcleos de las células que flotaban en el líquido amniótico. Estas células procedían del feto y, en las células de los tejidos de las hembras, la cromatina (el material nuclear afín al colorante) mostraba una condensación en forma de masa oscura: el corpúsculo de Barr. También pudimos determinar el grupo sanguíneo *ABO* del feto añadiendo células del líquido amniótico a soluciones de glóbulos rojos cuyo grupo sanguíneo se conocía, y agregando luego a la solución el anticuerpo para células de este grupo. En la solución en la que las dos poblaciones de células eran del mismo grupo (ambos tipos de células tenían moléculas marcadoras *ABO* dispuestas en su superficie) el anticuerpo daba lugar a una agrupación celular en la que participaban ambas poblaciones.

Las ventajas del diagnóstico prenatal son evidentes. Entre las numerosas anormalidades fetales que pueden detectarse ahora, muchas conducen a una enfermedad debilitante y otras ocasionan la muerte precoz. Algunas son causa de un retraso mental tan severo que impiden una vida normal. Actualmente no existe tratamiento para la mayoría de las anormalidades, pero si se detectan en una fase temprana de la gestación puede interrumpirse el embarazo. Incluso para los futuros padres que no consideren aceptable la interrupción del embarazo es importante conocer el diagnóstico, pues podrán entonces prepararse para el nacimiento de un niño afectado.

Se toman las muestras del líquido amniótico 16 a 17 semanas después del último período menstrual de la mujer embarazada. Es este el mejor momento para efectuar la punción, en parte porque el líquido amniótico ha alcanzado entonces un volumen de 175 a 225 mililitros, del que pueden llegar a extraerse hasta 25 mililitros sin colapsar el espacio líquido. El líquido tiene diversas fuentes: es segregado por el tracto respiratorio superior del feto, difunde a través de la piel fetal y a través de las membranas del saco amniótico. La fuente de líquido más constante es la orina excretada por el feto.

El líquido se extrae mejor inmediatamente después del examen de la imagen del feto mediante ultrasonografía. La esencia de esta técnica es que cuando vibraciones con frecuencias de 20.000 a varios millones de ciclos por segundo chocan con la interfase entre tejidos de diferentes densidades, se reflejan. Las reflexiones proporcionan una imagen sobre la que puede medirse el tamaño de diversas estructuras.

La sonografía permite diagnosticar anormalidades estructurales importantes. Las oscilaciones del sonograma revelan el latido cardíaco fetal varias semanas antes de que pueda detectarse con un estetoscopio corriente. La técnica por tanto hace posible el diagnóstico precoz de una muerte fetal. Con la ayuda de la sonografía se selecciona un lugar para la punción de modo que la aguja empleada para el procedimiento evite la cabeza del feto, y si es posible la placenta, y penetre en un acúmulo accesible de líquido. Un problema para el que la sonografía es especialmente valiosa es la detección de gemelos y otras gestaciones múltiples. Salvo raras excepciones, en una gestación múltiple cada feto tiene su propio saco amniótico, de modo que es necesario sacar una muestra del líquido de cada saco.

En cualquier caso el líquido siempre se extrae mediante una aguja larga y delgada que se inserta a través de la pared abdominal de la madre. Pequeñas muestras del líquido se utilizan para ensayos bioquímicos. El líquido restante se coloca en un frasco de cultivo de tejidos y se incuba para que las células vivas del líquido se multipliquen.

Una de las razones por las que se efectúan cultivos es que de éstos pueden aislarse un gran número de células durante la división, que es cuando el material genético está condensado en cromosomas. En otras fases el material genético es una maraña desorganizada que no puede analizarse. Los cromosomas se fotografían habitualmente bajo un microscopio de alta resolución y se identifican de acuerdo con su tamaño. En una técnica más moderna se tiñen con colorantes que se unen selectivamente a ciertas partes de los cromosomas, dando lugar a unos patrones de bandas característicos. La dotación cromosómica humana normal es de 46 cromosomas. Todos se encuentran en parejas excepto dos, los cromosomas sexuales, portadores de los genes que determinan que el hijo sea varón o hembra.

El grupo de enfermedades hereditarias que se diagnosticó en primer lugar mediante la amniocentesis fueron los trastornos ligados al sexo, que tienen su origen en un gen defectuoso que forma parte del cromosoma *X*. En los varones la presencia de este defecto da lugar a la manifestación de la enfermedad, pues la versión defectuosa del gen es la única copia de que dispone cada célula. En las hembras, no obstante, un cromosoma *X* defectuoso resulta enmascarado por un *X* normal. La enfermedad puede afectar a una hembra sólo si ambos cromosomas *X* son defectuosos, y esto es muy poco probable: requiere que la herencia genética de la madre incluya el gen defectuoso y que el padre padezca la enfermedad. Es aún más improbable dada la gravedad de las enfermedades, la mayoría de las cuales son invalidantes y acortan la vida notablemente. Dos ejemplos son la hemofilia, en la que la sangre carece de ciertas substancias que provocan su coagulación y por ende detienen la hemorragia, y la distrofia muscular de Duchenne, una debilidad muscular progresiva que ya en los primeros años de la adolescencia confina a su víctima a una silla de ruedas.

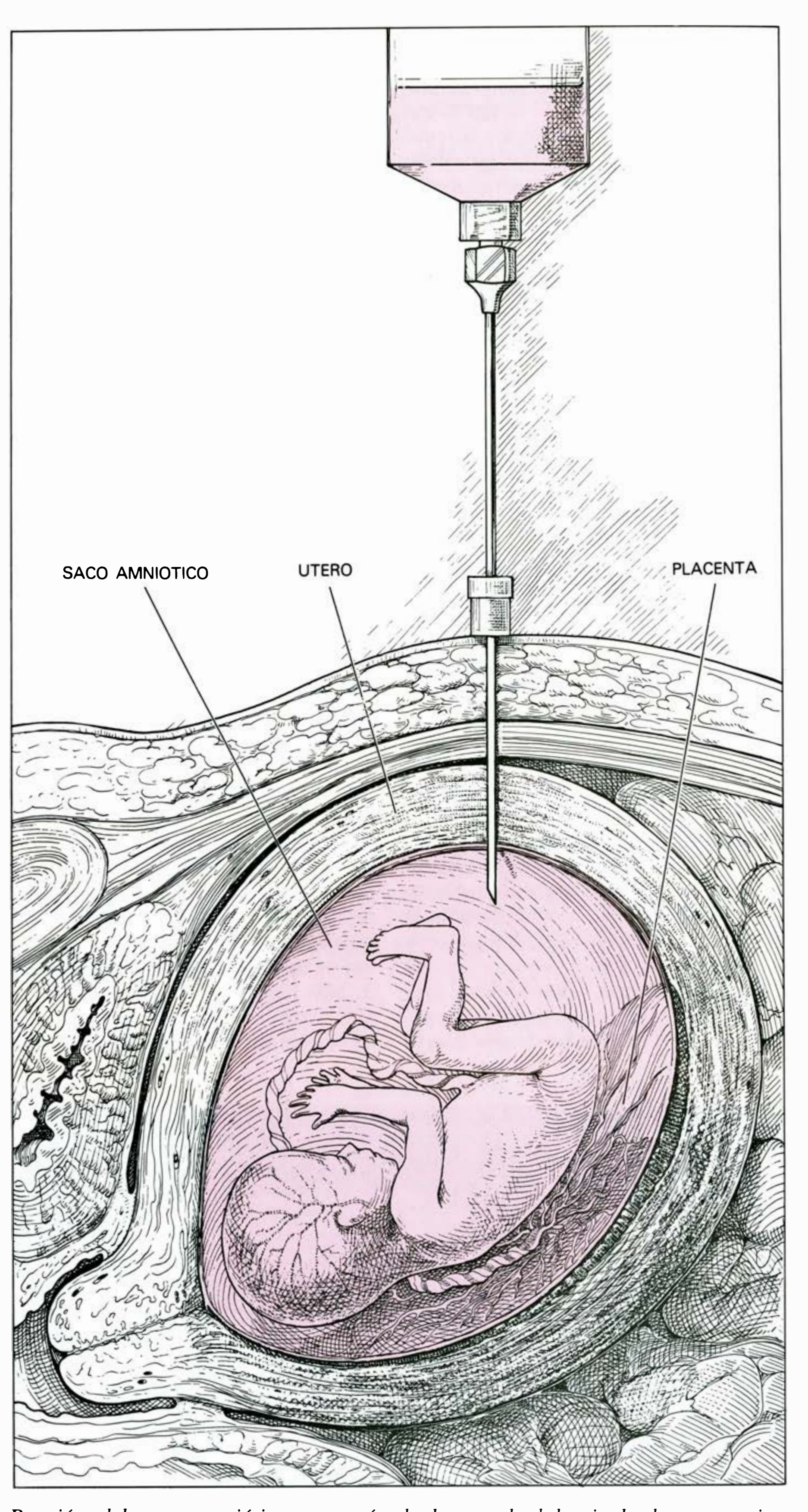
Por lo tanto, una mujer con un cromosoma *X* defectuoso es un portador que pasará el defecto ligado al sexo a la mitad de sus hijos. Por término medio, la mitad de los hijos varones sufrirán la enfermedad y la mitad de las hembras serán portadoras. Una mujer expuesta a dar a luz a un hijo con una enfermedad ligada al sexo suele estar al corriente del peligro, porque durante muchas generaciones algunos de sus antepasados varones se habrán visto afectados por la enfermedad. El examen de los cromosomas de las células del líquido amniótico revelará entonces el sexo del feto. Las células pueden teñirse para demostrar la presencia o ausencia del corpúsculo de Barr, o pueden cultivarse para identificar los cromosomas sexuales. Puesto que la enfermedad tiene un 50 por ciento de incidencia en el sexo expuesto (el varón), la indicación para amniocentesis es ciertamente muy poderosa. (Hay que añadir que los recientes esfuerzos encaminados a encontrar una correlación entre los trastornos ligados al sexo

y alguna característica del feto además del sexo, pueden finalmente facilitar el diagnóstico prenatal de los trastornos ligados al sexo en sí.)

Un segundo grupo de trastornos que pueden detectarse por amniocentesis son los causados por grandes anormalidades cromosómicas, normalmente un número anormal de cromosomas, pero a veces un defecto estructural importante en uno solo de ellos.

El más común de estos trastornos es el síndrome de Down, en el que cada célula del feto tiene tres de los cromosomas designados con el número 21, en vez de los dos normales. La anormalidad cromosómica se denomina trisomía del 21. Las consecuencias son retraso mental y malformaciones internas que a menudo incluyen enfermedad cardíaca congénita y que conducen a una muerte prematura. Típicamente la cabeza del niño es pequeña y aplanada y tiene características mongoloides, como son los ojos achinados. Otras anormalidades cromosómicas tienen características clínicas igualmente severas. Las hay compatibles con una vida normal, aunque son causa de esterilidad.

Anormalidades cromosómicas con consecuencias desastrosas para un feto pueden ser el resultado de un inocuo cambio en los cromosomas de uno u otro progenitor. Este cambio es una traslocación: el desplazamiento de un cromosoma, o parte de él, a un segundo cromosoma. Este defecto puede darse ya en todas las células del progenitor; en este caso el defecto existía en el óvulo fecundado del que derivaron todas las células. Por el contrario, el defecto puede existir en sólo algunas de las células del progenitor. Las células que presentan la anormalidad son las que proceden de una célula del embrión en la que la traslocación se dio por primera vez. En ambos casos la traslocación no causa ni pérdida ni ganancia de material cromosómico en ninguna célula del cuerpo, de ahí que el progenitor no manifieste enfermedad alguna. Las células germinales del progenitor sufren, no obstante, por meiosis, un proceso en el cual las células se dividen, pero sus cromosomas no se replican. En su lugar se distribuyen simplemente entre las células hijas, de modo que si un cromosoma se ha traslocado, una célula germinal puede recibir demasiado material cromosómico y otra puede recibir demasiado poco. Ambas condiciones son graves. Algunos fetos con células que incluyen 45 o 47 cromosomas son abortados espontáneamente. Otros llegan a tér-



Punción del saco amniótico, a través de la pared abdominal, de una mujer embarazada para extraer una porción del líquido que rodea al feto. Este procedimiento, denominado amniocentesis, proporciona muestras del líquido (y de las células que flotan en él) para el diagnóstico de enfermedades genéticas. El mejor momento para efectuar la amniocentesis es cuando el líquido ha alcanzado un volumen suficiente y su análisis permite decidir, en ciertos países, la interrupción del embarazo en caso de que se diagnostique alguna enfermedad grave

mino naciendo entonces con graves defectos.

Las trisomías pueden también presentarse por mutación. El riesgo de que ocurra aumenta al avanzar la edad materna, quizá porque todas las células germinales de los ovarios se formaron ya antes de que naciera la madre. Los gametos liberados del ovario durante los últimos años del período fértil materno han estado, por tanto, expuestas a mutágenos (agentes causantes de mutaciones) más tiempo que los ovulados antes.

Un tercer grupo de trastornos detectables por amniocentesis son las enfermedades causadas por aquellos defectos bioquímicos conocidos como errores congénitos del metabolismo. Normalmente el error congénito es la falta de un determinado enzima, consecuencia de un error en el gen que especifica su estructura. Aunque el gen incorrecto en sí no puede aún detectarse (en la mayoría de los casos el locus del gen, normal o anormal, aún no se conoce), el defecto se revela mediante ensayos del líquido amniótico o de sus células, que presentan tasas anormales de actividad enzimática, metabolitos anormales o metabolitos normales en concentraciones anormales. En unos pocos ejemplos el conocimiento de que una determinada enfermedad deriva de un determinado error congénito ha llevado al desarrollo de una terapia, pero la mayoría de errores congénitos acarrean enfermedades debilitadoras y graves que actualmente son incurables y presentan síntomas que no pueden aliviarse. Actualmente pueden diagnosticarse por amniocentesis unos 75 errores congénitos del metabolismo, bastante raro cada uno de ellos pero que en conjunto son la fuente de muchos padecimientos humanos.

Un ejemplo es la enfermedad de Tay-Sachs, en la que la sustancia grasa denominada gangliósido se acumula anormal y progresivamente en las células del sistema nervioso. Las consecuencias incluyen ceguera, demencia y parálisis. Normalmente el niño muere antes de alcanzar los cinco años. El error congénito que provoca la enfermedad de Tay-Sachs es una deficiencia del enzima hexosaminidasa *A*. Cuando se aplica calor al enzima, éste pierde su eficacia; con la ayuda de un ensayo basado en esta propiedad, un examen de células cutáneas, plasma sanguíneo o leucocitos tomados de los futuros padres revela si son portadores de los genes defectuosos. Los portadores son bastante raros en la población general, pero la incidencia es considerable entre los judíos procedentes de la Europa oriental. Allí la incidencia puede llegar a ser de uno entre 30.

Aunque no se conoce aún la localización precisa del gen defectuoso, no se halla en un cromosoma sexual. Es más, el gen defectuoso es recesivo, de modo que en los cromosomas del feto deben ser defectuosas las dos copias del gen para que se manifieste la enfermedad de Tay-Sachs. Dicho de otra manera, si los dos futuros padres son portadores, hay una posibilidad entre cuatro de que su hijo manifieste la enfermedad. En este caso es indicado efectuar una amniocentesis. Un ensayo de las células del líquido amniótico determinará entonces si el feto ha heredado la enfermedad.

El descubrimiento de David J. H. Brock y Roger Sutcliffe, del Western General Hospital de Edinburgh, de que la incidencia de ciertas malformaciones fetales está correlacionada con niveles elevados de la proteína denominada alfa-fetoproteína (AFP), tanto en el líquido amniótico como en la sangre de la madre, abrió el camino para el diagnóstico prenatal de un cuarto grupo de trastornos. La mayoría de ellos son defectos del tubo neural, la estructura del embrión de la que deriva el sistema nervioso central. Los defectos incluyen anencefalia, en la que falta, parcial o completamente, el cerebro del niño, además de la parte superior del cráneo, y espina bífida, un defecto que deja la médula espinal cubierta sólo por membranas o por piel y membranas y no por los huesos de la columna vertebral. También ciertas malformaciones en las que resultan afectados los riñones o el tracto gastrointestinal y la pared abdominal han sido asociadas con niveles anormalmente altos de AFP en el líquido amniótico.

La AFP se elabora en el hígado del feto, del cual es segregada a la circulación fetal en cantidades cada vez mayores hasta la vigésima semana del embarazo. De allí pasa el líquido amniótico; probablemente se excreta en la orina del feto, aunque también puede difundir a través de la piel. Pequeñas cantidades difunden luego a través de la placenta y del saco amniótico y entran en la circulación materna.

En la anencefalia se supone que hay un paso más: la AFP difunde a través de las delgadas paredes de los vasos sanguíneos en desarrollo del cerebro fetal, que está expuesto al líquido amniótico debido a la malformación congénita. La AFP no ocasiona por sí misma los defectos del tubo neural; se produce en cantidades normales en el hígado de un feto malformado. Lo que da importancia a la AFP es que es la única proteína conocida que existe en el feto y no se produce en los adultos. De modo que su presencia en la madre a niveles que parecen ser anormales puede ser un signo de diagnóstico.

No obstante, no hay una clara demarcación entre niveles maternos normales y anormales. Además, hay diversos factores que pueden inducir a una interpretación equivocada de la medición, como son la muerte no detectada del feto o un error en el cálculo de su edad. Esto último es una fuente de



Imagen del feto mediante ultrasonografía

error porque el nivel de AFP considerado normal varía con la edad del feto. Por consiguiente, para niveles maternos considerados límite, es importante practicar una sonografía para determinar la edad fetal y confirmar que el feto esté vivo antes de practicar la amniocentesis.

Se reconoció desde un principio que la inserción de una aguja en el saco amniótico y la extracción de una muestra de líquido amniótico podía asociarse a ciertos peligros, particularmente para el feto. Es por lo tanto importante valorar los riesgos de un trastorno genético frente a los peligros del procedimiento en cada caso individual.

Se conocen dos peligros principales: el aumento del riesgo de aborto espontáneo o ruptura del saco amniótico, que puede darse durante varias semanas después de la amniocentesis, y el riesgo de provocar en la madre una reacción inmunológica dirigida al feto. Es ésta una reacción en la que una madre *Rh* negativa (una mujer cuyos glóbulos rojos carecen de las moléculas de superficie denominadas factor rhesus) desarrolla anticuerpos que atacan los glóbulos rojos de un feto *Rh* positivo. La posibilidad surge porque la punción de la placenta con la aguja utilizada para la amniocentesis es a veces inevitable y puede significar que los glóbulos rojos fetales entren en la circulación materna. El peligro puede prevenirse, no obstante, inyectando globulina anti-*Rh* en la circulación materna de una madre *Rh* negativa. Esta substancia es un anticuerpo que destruye los glóbulos rojos *Rh* positivos que hayan entrado en la circulación materna antes de que alcancen a desencadenar la respuesta inmune. Si tanto la madre como el padre son *Rh* negativos no hay necesidad de llevar a cabo el tratamiento, porque en este caso el feto no puede ser *Rh* positivo y por tanto no hay reacción inmunológica.

El riesgo de la amniocentesis, pues, es esencialmente un aumento del peligro de aborto. Aumenta un riesgo natural de aborto, que se estima entre el 1 y el 2 por ciento en la fase del embarazo en que se practica la amniocentesis.

El aspecto más importante del futuro de la amniocentesis (y del diagnóstico prenatal general) es poner las técnicas presentes al alcance de todas las madres expuestas a dar a luz un niño anormal. Los recientes descubrimientos señalan el camino por el que cabe esperar el desarrollo del diagnóstico prenatal. Por un lado, la investigación continuada seguramente aumentará el número de errores congénitos del metabolismo detectables por amniocentesis. Al mismo tiempo puede esperarse que la lista de errores congénitos conocidos crezca al clarificarse los trastornos bioquímicos subyacentes a otras enfermedades genéticas.

Los errores congénitos, sin embargo, son raros. Una perspectiva más importante es el diagnóstico prenatal de enfermedades de elevada incidencia. Algunos ejemplos son la talasemia y la anemia falciforme, ambas debidas a anormalidades de la hemoglobina. Un enfoque de diagnóstico prenatal de patologías de la hemoglobina requiere la detección de la propia hemoglobina anormal en muestras de sangre fetal. Estas muestras se obtienen insertando en el saco amniótico una aguja más gruesa que la utilizada para la amniocentesis. A través de la aguja se inserta un delgado sistema fibroscópico que, a su vez, porta una fina y larga aguja. Puede así extraerse sangre de una arteria fetal gracias a que el médico ve realmente el vaso en la superficie de la placenta. Esta técnica, denominada fetoscopia, requiere gran habilidad y experiencia. A pesar de todo comporta un riesgo mucho mayor para el feto que la amniocentesis.

Este enfoque, no obstante, puede llegar a ser obsoleto. En 1978 Yuet Wai Kan y Andrée M. Dozy informaron de que la anemia falciforme podía diagnosticarse prenatalmente por medio de la amniocentesis. Su método se basaba en el uso de enzimas que rompían las cadenas de ADN en segmentos con secuencias nucleotídicas específicas. Las cadenas se tomaban del núcleo de las células del líquido amniótico. Algunas de las cadenas portaban el gen que codificaba para la hemoglobina. La rotura de estas cadenas por un enzima determinado producía fragmentos de un tamaño si el gen era normal y fragmentos de otro tamaño si el gen era defectuoso. La diferencia revelaba la existencia del defecto. Con el tiempo, técnicas similares posibilitarán el diagnóstico prenatal de muchas enfermedades genéticas para las que actualmente no se dispone de ningún método.

Los trastornos genéticos no pueden, hoy por hoy, curarse. Cuando el feto alcanza a ser visible por primera vez mediante sonografía, y por ende puede concebiblemente ser accesible a la ingeniería genética, consta ya de millones de células y en cada una están presentes copias de los genes y de cualesquiera defectos de los mismos. No obstante, ciertos errores congénitos del metabolismo tienen síntomas que pueden aliviarse después del nacimiento mediante medidas dietéticas o por la administración de fármacos. Es imaginable que un tratamiento de este tipo pueda aplicarse al feto. Además, es ya tópico en medicina que el primer paso en el tratamiento racional es un diagnóstico exacto y esto es lo que la amniocentesis proporciona en muchos casos. Los éxitos de la amniocentesis permiten esperar que algún día puedan tomarse medidas in utero para alterar la expresión de los genes y por consiguiente curar la enfermedad.

Crisis de identidad

El neutrino es la cosa más parecida a la nada. Se distingue de las restantes partículas elementales de materia principalmente por lo que no tiene y por lo que no hace. La teoría tradicional del neutrino lo describe como una partícula sin masa y sin carga eléctrica. Interactúa con la materia sólo a través de la más débil de todas las fuerzas básicas de la naturaleza, y estas interacciones son tan escasas que una corriente de neutrinos puede atravesar la Tierra sufriendo sólo una atenuación despreciable. La Tierra se mueve en una espesa sopa de neutrinos (se estima que existe una media de alrededor de 100 neutrinos por centímetro cúbico en todo el volumen del universo), pero son unas partículas tan efímeras que casi nunca dejan huella de su paso.

Esta descripción del neutrino tendrá probablemente que ser revisada pronto en lo referente a un detalle, aunque pequeño, crucial. Se está considerando seriamente la posibilidad de que al menos algunos tipos de neutrinos tengan masa. La cantidad de masa por partícula sería minúscula, pero en algunos aspectos importa menos el valor real que el mero hecho de que éste sea distinto de cero. Habida cuenta del elevado número de neutrinos, bastaría conferirles una masa pequeña para alterar drásticamente la estructura a gran escala del universo.

Las recientes especulaciones sobre la masa del neutrino guardan relación con el extraño fenómeno conocido por oscilaciones del neutrino. A lo largo de casi 20 años ha quedado bien sentado que existen por lo menos dos aromas, o clases, de neutrinos: uno puede aparecer sólo en asociación con un electrón y el otro aroma siempre se crea junto con un muon. Con posterioridad se ha añadido un tercero para acompañar a la partícula tau, recientemente descubierta, que es un hermano masivo del electrón y del muon. Hasta ahora, los

aromas se han tenido por inmutables, de suerte que un neutrino del tipo asociado al electrón nunca podría transformarse en un neutrino del tipo asociado al muon. Son precisamente estas transformaciones las que se verían posibilitadas por las oscilaciones del neutrino; en realidad, un único neutrino cambiaría continua y repetidamente de identidad al avanzar a lo largo de su trayectoria.

A fin de comprender el mecanismo de estas hipotéticas oscilaciones conviene describir el neutrino como una onda. La intensidad de la onda en cada punto nos da la probabilidad de encontrar en ese punto la partícula. En la teoría tradicional, cada aroma de neutrino tiene su propia onda independiente, y no puede haber superposición de las ondas asociadas con los diferentes tipos de neutrinos. Se introduce la posibilidad de oscilaciones al representar un único neutrino como una superposición de ondas de los tres tipos. El aroma del neutrino en cada punto viene entonces determinado por las amplitudes relativas de las tres componentes. Si las tres componentes tienen la misma amplitud, es igualmente probable que el neutrino tome cada uno de los tres aromas.

Incluso en la representación del neutrino como una superposición de ondas, la identidad de un neutrino dado permanecería estable indefinidamente con tal que las tres componentes se movieran a la misma velocidad. En este caso, las fases relativas de las tres ondas serían iguales en todos los puntos. Las oscilaciones sólo aparecerían si las tres componentes se movieran a distintas velocidades, de forma que las fases cambiaran continuamente y dominara primero una componente de la amplitud y luego otra. El fenómeno es similar a las «pulsaciones» que se escuchan al superponer dos ondas sonoras de frecuencia ligeramente distintas.

De acuerdo con la teoría especial de la relatividad, una partícula cuya masa sea exactamente cero deberá moverse siempre a la velocidad de la luz, en tanto que ninguna partícula dotada de masa alcanzará esa velocidad. De aquí se deduce que los neutrinos no podrían oscilar si todos ellos carecieran de masa. Recíprocamente, la observación de las oscilaciones del neutrino implicaría que por lo menos un tipo de neutrino tiene masa no nula. De las oscilaciones no sería posible deducir el valor real de la masa, pero a partir de su frecuencia o de su longitud de onda podríamos obtener la diferencia de masa entre los distintos tipos.

La posibilidad de las oscilaciones del neutrino fue propuesta por vez primera en 1963 (sólo un año después de haberse establecido la existencia de distintos tipos de neutrinos) por los fisicos japoneses M. Nakagawa, H. Okongi, S. Sakata y A. Toyoda. Cinco años más tarde, Bruno Pontecorvo, físico de origen italiano residente en la Unión Soviética, propuso por su cuenta una idea similar. Durante una década, la hipótesis siguió siendo un misterio, pero no se emprendieron esfuerzos conjuntos encaminados a la búsqueda de tales oscilaciones. El escepticismo sobre la existencia de las oscilaciones no era más que uno de los impedimentos; aun cuando existiesen, parecía dudoso que pudieran detectarse en un experimento real. Un reciente resurgimiento del interés sobre ese tema se debe en buena parte al trabajo procedente de las teorías de la «gran unificación» de las interacciones entre partículas elementales. En estas teorías, todavía en el terreno de la especulación, no sólo resulta plausible sino que es también probable que los neutrinos tengan una pequeña masa.

Frederick Reines, Henry W. Sobel y Elaine Pasierb, de la Universidad de California en Irvine, han dado cuenta de una observación que podría representar una señal de las oscilaciones del neutrino. Fue precisamente Reines, junto con Clyde L. Cowan, Jr., quien en la década de los 50 detectó por primera vez el neutrino; en el experimento definitivo se empleó un reactor nuclear de la Savannah River Plant, en Carolina del Sur. En la búsqueda de las oscilaciones del neutrino Reines y sus colaboradores han vuelto a la instalación de Savannah River. Para este último experimento se contaba con un detector cuyo medio sensible a los neutrinos lo constituían 268 kilogramos de óxido de deuterio, o agua pesada, excepcionalmente puro. Los instrumentos para detectar los productos de las alteraciones de los neutrinos se colocaron en el agua pesada, y se recubrió de plomo y cadmio la totalidad del aparato a fin de impedir la entrada de otras partículas que no fuesen los propios neutrinos.

La idea básica del experimento estaba en registrar simultáneamente dos clases distintas de interacción con neutrinos. Uno de estos procesos, denominado suceso de corriente neutra, puede iniciarlo cualquier aroma de neutrino; el otro proceso, conocido por suceso de corriente cargada, sólo puede ser desencadenado por un neutrino del tipo asociado al electrón (o, para ser más precisos, por un antineutrino del tipo asociado al electrón). En un reactor nuclear se generan neutrinos debido a la desintegración ß de los núcleos inestables resultantes de la fisión del uranio; se trata de antineutrinos exclusivamente del tipo asociado al electrón. Por tanto, en ausencia de oscilaciones, todos los neutrinos que alcancen el detector procedentes del reactor deben ser antineutrinos del tipo asociado al electrón; y los dos procesos medidos deben indicar el mismo flujo de partículas. Las oscilaciones cambiarían el aroma de algunos neutrinos, con el resultado de que se reduciría la proporción de sucesos de corriente cargada. Sin embargo, los sucesos de corriente neutra, que son inducidos por un igual por los tres tipos de neutrinos, no se verían afectados por las oscilaciones.

El experimento en cuestión se ha prolongado durante unos tres años, y todavía se están registrando datos. Reines y sus colaboradores interpretan que estos datos muestran que la proporción de sucesos de corriente cargada frente a los de corriente neutra es menos de la mitad de la proporción esperada. Si la explicación de la discrepancia radica en las oscilaciones, se deduce que el neutrino del tipo asociado al electrón y el del tipo asociado al muon difieren en masa (o en el equivalente energético de masa) en alrededor de un electronvolt. Un análisis de algunos datos más antiguos, registrados a una distancia del núcleo del reactor, es coherente con este resultado.

Reines ha destacado la persistencia de múltiples interrogantes relativas al experimento; sostiene que no se puede tomar como un descubrimiento de las oscilaciones del neutrino en tanto que no lo confirmen otras observaciones independientes. Las principales dificultades tienen su origen en la abundancia de neutrinos y en la rareza de sus interacciones. El reactor de Savannah River emite unos 10^{13} neutrinos por centímetro cuadrado y por segundo, pero éstos son tan reacios a interaccionar con la materia que tan sólo se detectaban por día unos 80 sucesos resultantes del flujo del reactor. Esta pequeña señal tenía que extraerse de entre un fondo de unos 400 sucesos diarios atribuidos a los neutrinos procedentes de otras fuentes.

Otros recientes experimentos también se han centrado en la cuestión de la masa del neutrino, y en conjunto sugieren que el asunto dista mucho de estar resuelto. Desde el pasado otoño un grupo de físicos procedentes del Instituto de Tecnología de California,

del Instituto Laue-Langevin de Grenoble y de la Universidad Técnica de Munich, ha estado midiendo el espectro de neutrinos emitido por un reactor de investigación sito en el Instituto Laue-Langevin. El rasgo característico por el que se detectan las interacciones con los neutrinos es algo diferente en este experimento, pero en principio la medida es sensible a cualquier fluctuación del tipo de neutrino. Felix H. Boehm, del Instituto de Tecnología de California, afirma que no se ha detectado prueba alguna de las oscilaciones.

En la Organización Europea para la Investigación Nuclear (CERN) se han efectuado tres experimentos completamente diferentes que, en visión retrospectiva, parecen mantener una estrecha relación con los estudios realizados en el reactor. Llamados experimentos de vaciado del haz, miden los porcentajes relativos de neutrinos del tipo asociado al electrón y del asociado al muon, presumiblemente creados en la desintegración de unas partículas de vida corta denominadas mesones *D*. Las primeras investigaciones sobre los mesones *D* habían mostrado que en estas desintegraciones se producen con igual probabilidad los dos tipos de neutrinos, pero el experimento de vaciado del haz registró un porcentaje doble de neutrinos del tipo asociado al muon. Este resultado causó gran sorpresa cuando fue comunicado, hace un año, y todavía hoy sigue rodeado de misterio. Las oscilaciones del neutrino son una posible explicación, pero no la única.

E. T. Tretyakov y sus colaboradores, del Instituto de Física Teórica y Experimental de Moscú, han anunciado recientemente otro enigmático resultado. El experimento de Tretyakov no se ocupaba de las oscilaciones del neutrino, sino que andaba detrás de otro fenómeno que podría dar una prueba del valor de la masa del neutrino. En este experimento se medía el espectro energético de los neutrinos emitidos en la desintegración del tritio, el isótopo radiactivo del hidrógeno. Si el neutrino del tipo asociado al electrón tuviese masa, este espectro presentaría un defecto de las energías más altas observadas; Tretyakov ha comunicado el hallazgo de tal defecto. Dicho resultado implica que el neutrino del tipo asociado al electrón tiene una masa del orden de los 10 electrovolt. Sin embargo, los detalles de este trabajo no son ampliamente conocidos por los físicos de fuera de la Unión Soviética.

En la búsqueda de las oscilaciones del neutrino se empeñaron también Lawrence R. Sulak, de la Universidad de Harvard, y sus colaboradores de Harvard y del Brookhaven National Laboratory. Los neutrinos se obtuvieron como un producto de desintegración secundario del haz del sincrotrón de protones de Brookhaven, que funcionó a una energía excepcionalmente baja a fin de asegurar que sólo se creasen neutrinos de tipo asociado al muon. La recogida de datos, que se prolongó durante varios meses, finalizó el pasado noviembre, pero todavía no está completo su análisis. Sulak estima que este experimento es diez veces más sensible que los anteriores.

Diez electronvolt representa una masa muy pequeña. La partícula más ligera conocida que tenga masa distinta de cero es el electrón, y su masa es de 511.000 electronvolt. Sin embargo, incluso la más pequeña masa medible de los neutrinos tendría enormes repercusiones no sólo en la física de partículas elementales sino también en astrofísica y en cosmología. Así, las oscilaciones (de existir) podrían resolver una controversia sobre la química nuclear del Sol que viene de antiguo. Durante más de una década Raymond Davis, Jr., y sus colaboradores de Brookhaven han estado trabajando con un gigantesco detector de neutrinos hundido en el pozo de una mina a fin de medir el flujo de electrones procedentes del Sol. El flujo observado ha resultado alcanzar sólo un tercio del valor que se esperaba. En el experimento de Davis, el detector empleado es sensible sólo a los antineutrinos del tipo asociado al electrón, por lo que las oscilaciones podrían ofrecer una explicación alternativa. Si los antineutrinos del tipo asociado al electrón oscilasen para dar iguales proporciones de los tres tipos de neutrinos antes de alcanzar la Tierra, el detector tan sólo registraría un tercio del valor esperado.

Puesto que los neutrinos son tan abundantes, el descubrimiento de que poseen incluso una pequeña masa podría más que duplicar la masa total del universo. Esta duplicación alteraría a su vez substancialmente las interacciones gravitatorias de la materia en las galaxias. Las observaciones han sugerido que la mayoría de las galaxias son miembros de cúmulos. Sin embargo, es curioso que las masas aparentes de las galaxias, estimadas a partir del número de estrellas y de la cantidad de polvo y gas presentes, parece ser unas 10 veces demasiado pequeña para mantener unidos los cúmulos. Un mar de neutrinos con una masa de 10 electronvolt extendido por todo el espacio podría explicar la cohesión de los cúmulos. A una escala todavía mayor, la masa de los neutrinos influiría en la evolución global del universo. Las estimaciones de la masa visible de las galaxias y de la materia intergaláctica sugieren que el universo tiene una estructura «abierta», significando esto que la expansión observada continuará indefinidamente. Un valor suficientemente grande de la masa de los neutrinos podría llegar a frenar la larga expansión y conducir a un gran colapso.

Las consecuencias del valor de la masa del neutrino seguirán siendo una especulación inútil hasta que quede bien establecido que su masa no es nula. Es de esperar que dentro de algunos meses se anuncie el resultado de la investigación de Sulak, lo cual podría esclarecer en parte el problema. Con algunos experimentos más, que ahora están en preparación, se confía que esta cuestión quede resuelta.

Reines y sus colaboradores están instalando un nuevo detector en el reactor de Savannah River. El detector y su revestimiento se montarán sobre ruedas, de suerte que las medidas podrán efectuarse a diferentes distancias del núcleo del reactor. Si la longitud de onda de las oscilaciones de los neutrinos fuera sólo de algunos metros, como indican los primeros resultados de Reines, sería posible observar picos y valles en las componentes de los neutrinos, simplemente con mover el aparato. De forma similar, el grupo que en la actualidad trabaja en el Instituto Laue-Langevin se traslada este mismo verano a otro reactor más potente y también realizará medidas a diferentes distancias, y los experimentos de vaciado del haz realizados en el CERN se van a repetir empleando detectores más sofisticados en nuevas posiciones.

Un grupo de investigadores procedentes de la Universidad de Pennsylvania, de la Brown University, de Brookhaven y del Japón están instalando en Brookhaven un detector para una nueva investigación. El medio de detección consistirá en 150 toneladas de un líquido que centellea cuando lo atraviesa una partícula cargada. Sulak y sus colaboradores de la Universidad de Michigan han propuesto un plan todavía más ambicioso para el haz de neutrinos de Brookhaven, consistentes en enterrar siete detectores a lo largo del haz subterráneo a distancias de hasta un kilómetro de la fuente. Cada detector estaría formado por un recipiente con unas 200 toneladas de agua ultrapura.

Fisiología de las plaquetas sanguíneas

Estos corpúsculos juegan un papel esencial en la detención del flujo sanguíneo en las heridas. Sus interacciones con sustancias del plasma sanguíneo y de los tejidos desempeñan múltiples funciones en condiciones fisiológicas y patológicas

Marjorie B. Zucker

En una extensión de sangre vista al microscopio óptico pueden observarse las células rojas discoidales, una gran variedad de células blancas llamativamente teñidas y, quizá menos impresionantes, unos pequeños fragmentos de materia granular, las plaquetas: los elementos formes más pequeños de la sangre. Las células rojas (o glóbulos rojos) transportan oxígeno; las células blancas (o glóbulos blancos) poseen diversas funciones que protegen al organismo frente a sustancias extrañas y eliminan detritos celulares. Las plaquetas desempeñan un papel esencial en la hemostasia, la detención del flujo de sangre. Se agrupan y forman un tapón que temporalmente cierra cualquier solución de continuidad en la pared del vaso. Sustancias intraplaquetarias o de la superficie de las plaquetas, junto a otras sustancias del plasma sanguíneo o de la pared de los vasos, contribuyen a la formación de un tapón permanente o, en diferentes circunstancias, a la formación de un trombo que puede obstruir la circulación y ser la causa de un ataque cardíaco o de una apoplejía.

La existencia y algunas de las funciones de las células rojas y blancas se conocen desde que empezó a utilizarse el microscopio óptico; sin embargo, la existencia de las plaquetas no se estableció definitivamente hasta finales del siglo XIX. Pronto se supo que estaban implicadas en la hemostasia y coagulación, pero su papel en dichos procesos no se descifró hasta finales de la década de los cuarenta y en la década de los cincuenta [véase "Blood Platelets", por Marjorie B. Zucker; SCIENTIFIC AMERICAN, febrero de 1961]. Desde entonces se han descubierto muchos detalles sobre la fisiología y bioquímica de la función plaquetaria, así como sobre los mecanismos implicados en procesos patológicos en los que intervienen las plaquetas, ya sea porque no llevan a cabo su función, o porque la ejecutan demasiado bien, aunque en un sitio inadecuado.

Las plaquetas y su contenido

Si la sangre recién extraída se mezcla con un anticoagulante para evitar la coagulación y se deja en reposo para que sedimente, algo más de la mitad de su volumen aparece en la parte superior como un líquido amarillento rico en proteínas. Se trata del plasma sanguíneo. La mayor parte del resto está constituido por las células rojas; las células blancas y las plaquetas representan solamente del 1 al 2 por ciento del volumen total de la sangre. Cada milímetro cúbico de sangre contiene alrededor de un cuarto de millón de plaquetas (en comparación con sólo 7000 células blancas), de modo que hay alrededor de un billón de plaquetas en la sangre de una mujer cuyo volumen de sangre fuera de cuatro litros. Las plaquetas no son células sino fragmentos de células gigantes de la médula ósea denominadas megacariocitos. A medida que el megacariocito madura, su citoplasma (el material que envuelve al núcleo) se fragmenta para formar varios miles de plaquetas, corpúsculos de morfología más o menos discoidal con un diámetro entre un medio y un tercio del correspondiente al de un glóbulo rojo, pero con un volumen que representa sólo una treceava parte del volumen de esta célula.

Al carecer de núcleo, las plaquetas no contienen ADN y muestran una escasa capacidad de síntesis de proteínas. Una vez liberadas de la médula ósea envejecen lentamente en la circulación y mueren al cabo de 10 días, si no han sido consumidas antes en acto de servicio. No obstante, las plaquetas muestran un metabolismo muy activo que corre a cargo de enzimas del citoplasma o de los orgánulos intracelulares denominados mitocondrias, que proporcionan la energía requerida para sus funciones.

Las plaquetas observadas con el microscopio óptico en una extensión teñida de sangre presentan un aspecto granuloso; el microscopio electrónico revela una gran variedad de orgánulos bien definidos. Una banda circular de largos microtúbulos situada en la proximidad de la membrana externa sirve de andamio estructural para mantener la forma aproximadamente discoidal de la plaqueta. Existen numerosas vesículas abiertas, o canales membranosos, algunos de los cuales están conectados entre sí y con la superficie de la plaqueta de forma similar a los canales de una esponja. Constituyen el denominado sistema conectado con la superficie. Además de las mitocondrias y de pequeños acúmulos de glucógeno (la forma de almacenamiento de la glucosa) se aprecian otros tipos de gránulos. Se trata de vesículas cuyo contenido se segrega al estimular las plaquetas.

Los distintos tipos de gránulos pueden separarse rompiendo cuidadosamente las membranas externas de las plaquetas y sometiendo a centrifugación en gradientes de sacarosa la suspensión de orgánulos resultante. Un gradiente de sacarosa consiste en un tubo de centrífuga que contiene una disolución de sacarosa cuya concentración aumenta progresivamente a medida que nos aproximamos al fondo del tubo. Los distintos orgánulos se desplazan a través del gradiente hasta que alcanzan un nivel en el que la densidad de la sacarosa coincide con la densidad del gránulo; en este punto, el gránulo se detiene. Los gránulos más densos aparecen en el fondo del tubo. Los fragmentos de membrana externa

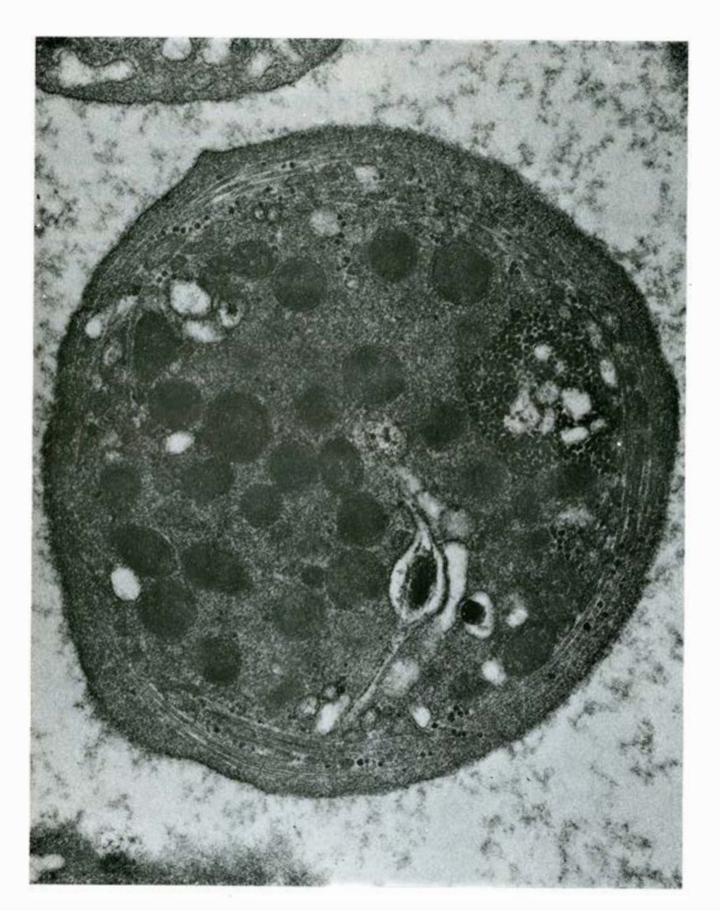
están situados cerca de la superficie, y otros componentes quedan dispuestos en bandas definidas a lo largo del tubo. La composición de cada banda puede determinarse con el microscopio electrónico y mediante análisis químico.

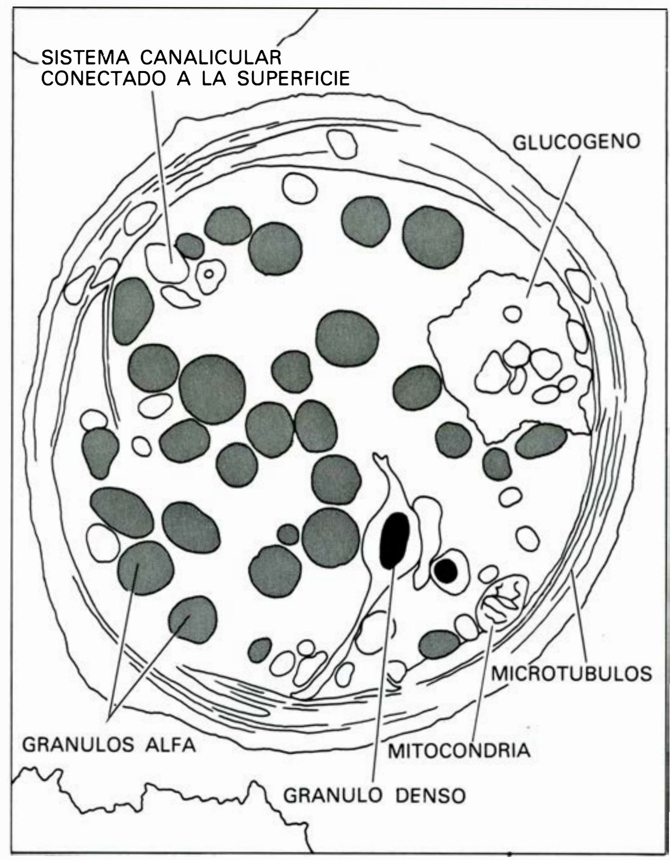
Los gránulos que muestran la mayor densidad en las micrografías son también los más densos en el gradiente. Estos gránulos contienen una mezcla concentrada de serotonina, calcio y los nucleótidos difosfato de adenosina (ADP) y trifosfato de adenosina (ATP). Estos últimos no intervienen en la actividad metabólica, a diferencia del ADP y ATP del citoplasma de la mayoría de las células y de las plaquetas; no establecen intercambio con los nucleótidos del citoplasma, del mismo modo que el calcio de los gránulos no se intercambia con el calcio citoplásmico. Tampoco la serotonina de los gránulos (que se sintetiza en células de la pared intestinal, circula por la sangre y es captada por la membrana externa de las plaquetas) se intercambia con el neurotransmisor serotonina del cerebro.

Los gránulos alfa, más ligeros y numerosos que los gránulos densos, contienen distintas proteínas, probablemente sintetizadas en los megacariocitos. Algunas de ellas sólo se encuentran en las plaquetas. Cabe citar el factor plaquetario 4, que contrarresta al agente anticoagulante heparina, y un potente factor de crecimiento. Otras proteínas son semejantes o quizás idénticas a proteínas del plasma implicadas

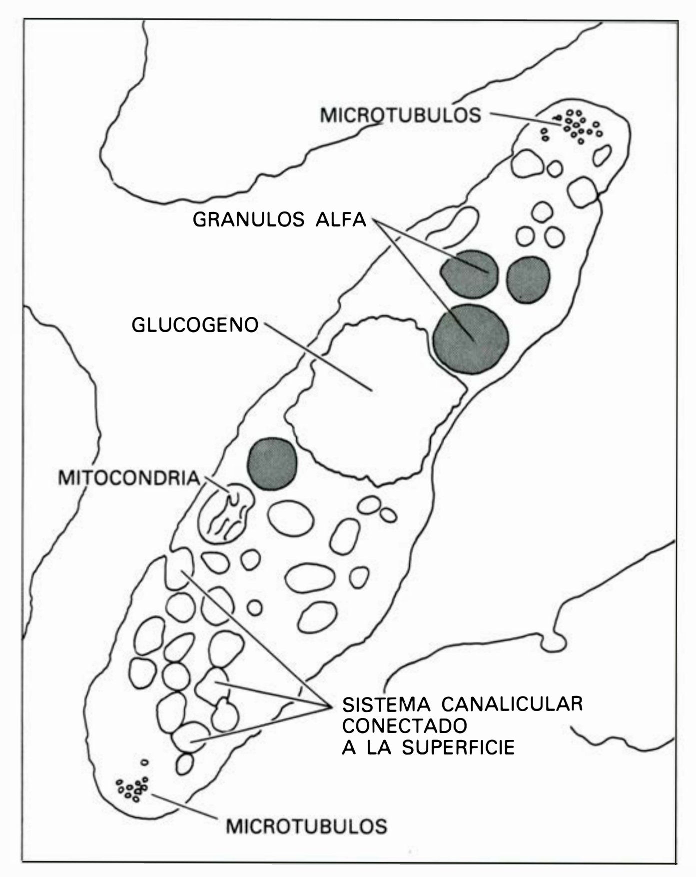


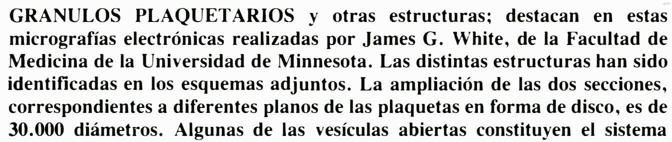
PLAQUETA BLOQUEANDO UNA ABERTURA en un capilar. La micrografía electrónica fue obtenida por Hans R. Baumgartner, de F. Hoffmann-La Roche y Co. de Basilea. La sección transversal muestra cuatro glóbulos rojos y una plaqueta con una ampliación de 13.000 diámetros. El capilar se ha dilatado artificialmente y dos de las células endoteliales que lo recubren han quedado separadas (*abajo*), dejando expuesto el subendotelio a la sangre. Las plaquetas, que no se adhieren a la capa endotelial del vaso, sí lo hacen al colágeno, un constituyente del subendotelio. En este caso se ha adherido una sola plaqueta, que cierra la abertura. En ausencia de plaquetas, la extravasación de sangre entre células endoteliales de vasos diminutos produce unas hemorragias puntiformes en la piel, denominadas petequias. Las soluciones de continuidad en vasos de gran tamaño con paredes mucho más gruesas que las de los capilares se ocluyen mediante la agregación de una considerable cantidad de plaquetas (*véase la ilustración de la página 52*).











GRANULOS PLAQUETARIOS y otras estructuras; destacan en estas micrografías electrónicas realizadas por James G. White, de la Facultad de Medicina de la Universidad de Minnesota. Las distintas estructuras han sido identificadas en los esquemas adjuntos. La ampliación de las dos secciones, correspondientes a diferentes planos de las plaquetas en forma de disco, es de 30.000 diámetros. Algunas de las vesículas abiertas constituyen el sistema conectado con la superficie. Los microtúbulos integran el andamiaje estructural del disco plaquetario, en tanto que las mitocondrias son orgánulos de transducción energética. El glucógeno, una forma de almacenamiento de la glucosa, es una fuente de energía. Los gránulos densos y los gránulos alfa contienen diversas sustancias que se segregan al recibir las plaquetas un estímulo y que promueven la formación de los tapones plaquetarios.

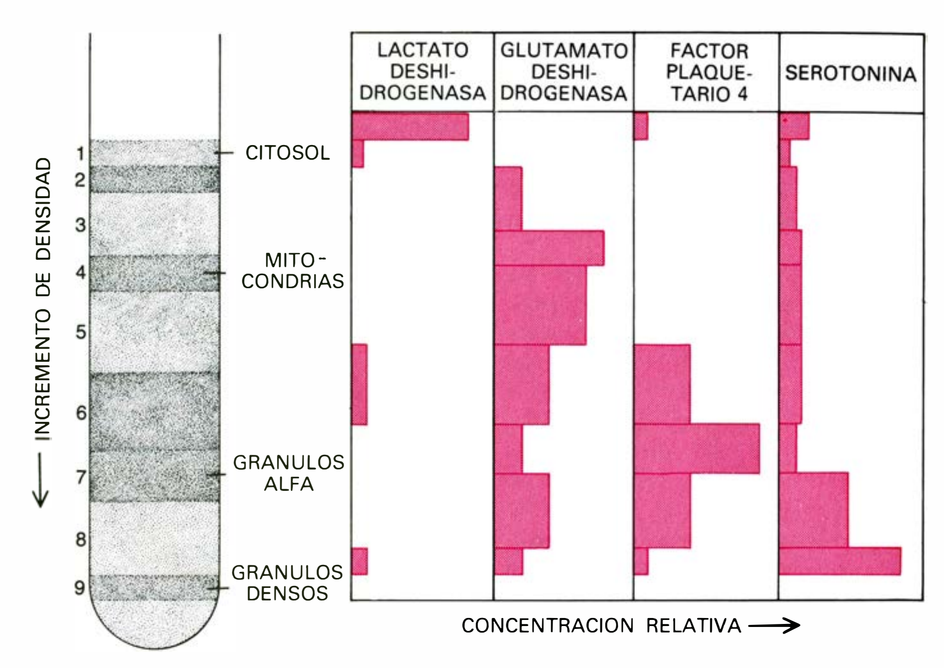
en el proceso de coagulación, tales como el fibrinógeno y el factor V. Las plaquetas poseen un tercer tipo de gránulo, los lisosomas, presentes en la mayoría de las células; poseen enzimas que hidrolizan proteínas y carbohidratos complejos.

Hemostasia

Las plaquetas poseen una mayor cantidad de la proteína contráctil actomiosina que cualquier otra célula no muscular, y se contraen de forma parecida a como lo hacen los músculos. La actividad contráctil de las plaquetas puede ponerse de manifiesto en el tubo de ensayo en el proceso de retracción del coágulo. Cuando se coloca en un tubo de ensayo sangre recién extraída, coagula en unos 10 minutos, formando una masa gelatinosa de color rojo. Al cabo de una hora, aproximadamente, el coágulo se retrae, disminuyendo de tamaño, y expulsa el suero, un líquido claro de color amarillo. La fracción de sangre que solidifica para formar el coágulo es la proteína plasmática llamada fibrinógeno. El fibrinógeno se transforma por la acción del enzima trombina en hebras de fibrina, que forman una malla en la que quedan atrapadas las plaquetas y otras células sanguíneas. La mayor parte de estas células son glóbulos rojos responsables del color del coágulo. La trombina estimula también la contracción de las plaquetas que se han adherido a las hebras de fibrina; la contracción aproxima las hebras y expulsa el suero (plasma desprovisto de fibrinógeno). El proceso de retracción del coágulo, aunque espectacular en el tubo de ensayo, no tiene contrapartida en el organismo. Las soluciones de continuidad de los vasos sanguíneos no se ocluyen por fibrina y glóbulos rojos, sino por una masa de plaquetas desprovista de eritrocitos.

Los vasos sanguíneos intactos poseen un fino recubrimiento de células endoteliales que evita el contacto de las plaquetas con otros elementos de la pared del vaso. Al seccionar un vaso, la sangre fluye a través del corte y uno de los constituyentes de la pared del vaso, el colágeno, la proteína fibrosa insoluble que constituye la mayoría del tejido conectivo del organismo, interacciona con las plaquetas. Estas se adhieren rápidamente al colágeno y a otras plaquetas que atraviesan la pared seccionada formándose en un minuto un tapón hemostático. A continuación, entra en juego el proceso contráctil que se observaba en la retracción del coágulo en el tubo de ensayo, que compacta la masa plaquetaria y expulsa las sustancias segregadas por las plaquetas.

Se pone en marcha, además, un segundo proceso. La pared del vaso seccionado (al igual que otros tejidos lesionados) contiene un factor tisular que inicia la coagulación sanguínea. La exposición de la sangre a dicho factor dispara una compleja serie de reacciones en virtud de las cuales una proteína plasmática, la protrombina, se convierte en trombina. Al igual que vimos en el tubo de ensayo, la trombina, a su vez, transforma el fibrinógeno en fibrina. La fibrina formada en la pared del vaso lesionado sirve de soporte y refuerza el tapón plaquetario, y, al mismo tiempo, solidifica la sangre que permanece en la herida. Vamos a ocuparnos ahora de la función plaquetaria y el tapón plaquetario sin tratar del proceso global de la coagulación.



GRANULOS Y OTRAS FRACCIONES PLAQUETARIAS separados por centrifugación en un gradiente de sacarosa, en el que el material de densidad creciente se dispone en forma de bandas en niveles progresivamente inferiores del tubo *(izquierda)*. Se indica el componente principal de cuatro de las bandas, determinado por microscopía electrónica. Las cuatro columnas de barras horizontales muestran la concentración relativa en cada banda de cuatro constituyentes químicos de las plaquetas. La lactato deshidrogenasa, un enzima del citoplasma, se encuentra en el líquido sobrenadante; la glutamato deshidrogenasa es un enzima mitocondrial. El factor plaquetario 4 se encuentra en los gránulos alfa, y la serotonina en los gránulos densos. Los datos correspondientes a estudios de M. J. Brockman, N. P. Westmoreland, P. Cohen y K. L. Kaplan. El contenido plaquetario es fundamental en la hemostasia.

Agregación plaquetaria

La trombina tiene una doble acción en la hemostasia. Además de convertir el fibrinógeno en fibrina, hace que las plaquetas se vuelvan pegajosas. La exposición de las plaquetas al colágeno y a la trombina determina su agregación y la formación de un tapón que ocluye la solución de continuidad en la pared del vaso sanguíneo. El proceso de la agregación plaquetaria puede investigarse con un simple instrumento denominado agregómetro. Un plasma rico en plaquetas puede prepararse centrifugando a pocas revoluciones sangre en presencia de un anticoagulante, con lo que sedimentan los glóbulos rojos y blancos y permanecen en suspensión las plaquetas, elementos más ligeros. Si el plasma rico en plaquetas se coloca en un tubo transparente y se somete a agitación, cuando un haz de luz atraviesa el tubo, una nube difusa de diminutas plaquetas dispersa la luz, evitando que la mayoría del haz alcance la fotocélula situada en la cara opuesta. Al añadir un agente agregante, las plaquetas se agrupan, permitiendo que pase más luz a través de la suspensión. Un trazado de la cantidad de luz que alcanza la fotocélula permite registrar de forma continua el curso de la agregación plaquetaria.

Si se añade una suspensión de pequeñas partículas de colágeno a un plasma rico en plaquetas, no se observa efecto alguno durante un minuto, iniciándose luego la agregación. El fenómeno inicial consiste en una interacción de ciertos puntos específicos de la molécula de colágeno y glicoproteínas específicas (proteínas que contienen azúcar) de la superficie de las plaquetas. Como consecuencia de la interacción se produce la adhesión de las plaquetas al colágeno. Este proceso pone en marcha una serie de mecanismos, no muy

bien conocidos, en la membrana de las plaquetas que culminan con la secreción del contenido de los gránulos densos y de los gránulos alfa. El contenido de los gránulos se encuentra ahora en el plasma y el examen de las plaquetas muestra que han perdido sus gránulos.

Uno de los componentes de los gránulos, el ADP, es un potente agente de agregación. Añadiendo ADP purificado a una concentración muy baja, del orden de 1 micromolar, se inicia la agregación al cabo de segundos. La concentración de ADP liberado en una suspensión plaquetaria a la que se ha añadido colágeno puede alcanzar un valor comprendido entre 5 y 10 micromolar. La secreción de ADP sería en gran parte (y quizás enteramente) responsable de la agregación causada por el colágeno: el tiempo requerido para la secreción explicaría el período de latencia existente entre la adición del colágeno y el comienzo de la agregación.

La respuesta inicial de las plaquetas estimuladas por ADP es un cambio de forma. Largas extensiones se proyectan desde la superficie de las plaquetas y el disco se hace más grueso hasta que adquiere el aspecto de una esfera irregular y erizada. Al mismo tiempo, aparecen en la superficie de las plaquetas receptores específicos para el fibrinógeno; su presencia puede demostrarse por la capacidad de unirse al fibrinógeno marcado radiactivamente que muestran las plaquetas estimuladas por ADP. El ADP conduce a la agregación únicamente en presencia de fibrinógeno, que se encuentra normalmente en el plasma. Ciertas manipulaciones del medio no sólo evitan la agregación dependiente del ADP, sino también la unión del fibrinógeno. Esto sugiere que el fibrinógeno actúa como una especie de cemento de unión entre las plaquetas. La molécula de fibrinógeno está capacitada para tal función por tratarse de un dímero (una molécula doble); cada molécula podría unirse a una plaqueta diferente. El ADP no altera la membrana plaquetaria de forma permanente. Cuando desaparece del medio, las plaquetas liberan el fibrinógeno que se había unido a su superficie, se desagregan y pueden estimularse de nuevo.

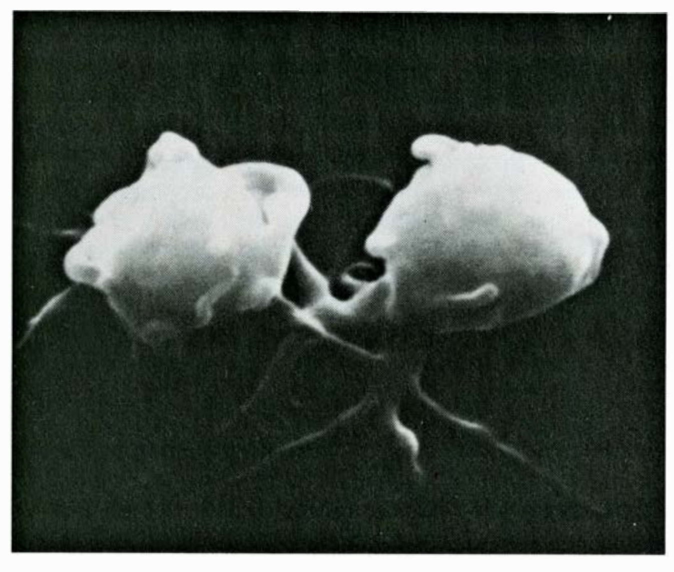
Secreción

La trombina estimula la agregación directamente, como el ADP, y también estimula la secreción. A bajas concentraciones la trombina añadida a un plasma rico en plaquetas produce la agregación en dos fases: una agregación moderada, que ocurre al cabo de unos diez segundos, y otra más intensa, que tiene lugar 1 minuto después. La primera agregación se debe a la trombina misma, mientras que la segunda depende del ADP segregado. La trombina estimula las plaquetas al interaccionar con una glicoproteína de la membrana plaquetaria, a la que hidroliza. También el fibrinógeno sería en este caso el cemento de unión, pero en la estimulación por trombina el fibrinógeno procede de la secreción del contenido de gránulos plaquetarios y no es necesario que se halle previamente en el medio.

Aparentemente son dos las vías que conducen a la secreción, habiéndose descifrado en la actualidad solamente una de ellas. La estimulación de las plaquetas por adhesión al colágeno o por trombina libera ácido araquidónico, un ácido graso presente en la membrana plaquetaria, que es rápidamente transformado en compuestos denominados endoperóxidos, los cuales, a su vez, se convierten en tromboxano A_2, que constituye el estímulo determinante de la secreción. La aspirina y otras drogas antiinflamatorias pueden bloquear la conversión del ácido araquidónico en tromboxano A_2. La estimulación plaquetaria en una persona que ha ingerido aspirina produce una respuesta de agregación anormal. El colágeno induce mucha menos agregación y bajas concentraciones de trombina producen únicamente la fase inicial o directa. La existencia de un segundo mecanismo de secreción se deduce a partir del hecho siguiente: una estimulación más fuerte, por colágeno o mediante trombina, de plaquetas tratadas con aspirina, determina cierta secreción incluso cuando no se forma tromboxano A_2.

La secreción y la agregación plaquetarias actúan conjuntamente en la hemostasia. La adhesión de las plaquetas al colágeno en el vaso lesionado determina la secreción de ADP, el cual estimula la agregación. Otras plaquetas se unen a las ya adheridas al colágeno. La serotonina segregada a partir de los gránulos densos es un agente agregante débil y también un vasoconstrictor que disminuye el calibre de los vasos localmente, tanto si están seccionados como intactos, y colabora así en el proceso de cohibir la hemorragia. El fibrinógeno, un cofactor necesario para la agrega-





CAMBIO DE FORMA de las plaquetas al ser expuestas al ADP, nucleótido liberado a partir de los gránulos densos. Las plaquetas frescas aparecen como discos aplanados de contornos lisos en la micrografía (*izquierda*) realizada por White. Tras el tratamiento con ADP, proyectan largos procesos desde su superficie, la plaqueta se hincha y se transforma en una esfera "erizada"; las prolongaciones se adhieren a una superficie (*derecha*) o a otras plaquetas.

ción, está presente en el plasma y se segrega también a partir de los gránulos alfa; parte del fibrinógeno se convierte en fibrina por la acción de la trombina. El factor V, que ayuda en la conversión de protrombina en trombina, está presente en el plasma, pero probablemente se segrega a partir de los gránulos alfa en una forma más activa; este factor se adhiere a la superficie de las plaquetas y determina la producción local de una mayor cantidad de trombina. Sustancias de la pared vascular, del plasma y de las plaquetas (del interior y de la superficie) actúan conjuntamente en la formación del tapón que detiene la hemorragia.

Función plaquetaria anormal

¿Cuál es el papel de los distintos procesos plaquetarios en la prevención de las hemorragias? Una forma de averiguarlo es determinar el "tiempo de sangría" en individuos normales a los que se han administrado drogas que alteran la función plaquetaria o en personas que muestran alteraciones plaquetarias derivadas de enfermedades congénitas. La prueba se realiza elevando la presión de la sangre en los vasos de pequeño calibre (aplicando un manguito en un brazo) y practicando cortes diminutos en el antebrazo. Las gotas de sangre se limpian cada 30 segundos sin hacer presión sobre el corte. En condiciones normales, la hemorragia se detiene al cabo de 6 minutos, pero si la función plaquetaria es anómala, la hemorragia continúa hasta que es detenida por aplicación directa de presión.

La necesidad de las plaquetas resulta evidente: los pacientes con un número escaso de plaquetas, trombocitopenia (trombocito es el antiguo término aplicado a las plaquetas), tienen un tiempo de sangría prolongado. Estos pacientes sangran excesivamente después de pequeñas lesiones, o incluso espontáneamente, en particular en las membranas mucosas. La trombocitopenia puede resultar de una producción deficitaria de plaquetas en la médula ósea, ya sea de origen congénito, o producida por una leucemia, una anemia aplástica o por la quimioterapia del cáncer. En algunas ocasiones, sin embargo, un individuo puede producir anticuerpos frente a sus propias plaquetas, siendo éstas destruidas casi tan rápidamente como se producen. Puesto que la enfermedad es idiopática (de causa desconodida) y el síntoma más destacado es la púrpura, pequeños puntos hemorrágicos (petequias) y moraduras en la piel, este proceso recibe el nombre de púrpura trombocitopénica idiopática o PTI.

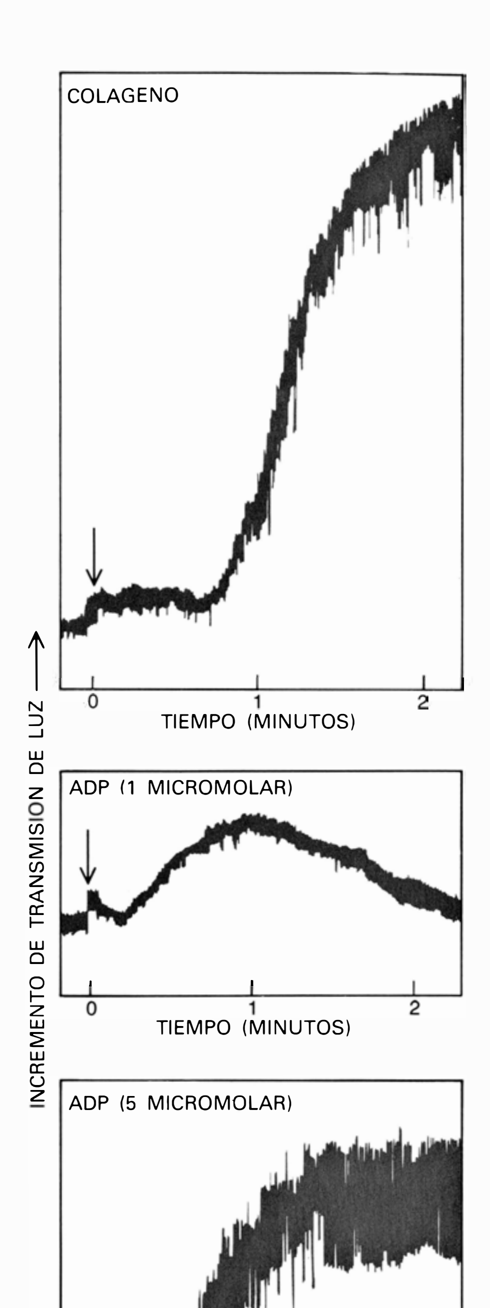
Ciertos albinos muestran una ausencia congénita de gránulos densos en sus plaquetas, lo que se refleja también en un prolongado tiempo de sangría. La ingestión de aspirina, que interrumpe la vía de secreción, prolonga ligeramente el tiempo de sangría, si bien no produce hemorragias excesivas en individuos normales. Ambas observaciones indican que la secreción plaquetaria es importante para la hemostasia. Pacientes con una rara enfermedad hemorrágica denominada tromboastenia tienen un largo tiempo de sangría; sus plaquetas no pueden agregar en absoluto en presencia de colágeno, trombina o ADP, y son incapaces de unirse al fibrinógeno. Personas con un déficit congénito de fibrinógeno, una enfermedad muy rara, muestran también un tiempo prolongado de sangría, por lo que el papel esencial del fibrinógeno en la agregación resulta evidente.

Sorprendentemente, pacientes con ciertas alteraciones de coagulación como la hemofilia no muestran un tiempo de sangría prolongado, aunque sufran frecuentes e importantes hemorragias espontáneamente o como consecuencia de traumas mínimos, una extracción dentaria o un pequeño golpe, por ejemplo. La hemofilia se produce por la ausencia del denominado factor antihemofílico, una proteína plasmática que junto a otros factores resulta esencial para la transformación de protrombina en trombina. Puesto que el mecanismo de formación de trombina dependiente del factor tisular no requiere factor antihemofílico, se produce suficiente trombina para detener la hemorragia causada por las diminutas lesiones de la prueba del tiempo de sangría.

Enfermedad de von Willebrand

Otros pacientes con tiempo de sangría prolongado padecen una alteración hereditaria conocida como enfermedad de von Willebrand. En estos pacientes, la actividad de dos proteínas plasmáticas, el factor antihemofílico y el factor de von Willebrand (que normalmente van unidos), es baja. El último factor es necesario para un tiempo de sangría normal. Su función resulta difícil de comprender puesto que las plaquetas de estos enfermos se agregan normalmente en el tubo de ensayo en presencia de ADP, colágeno y trombina.

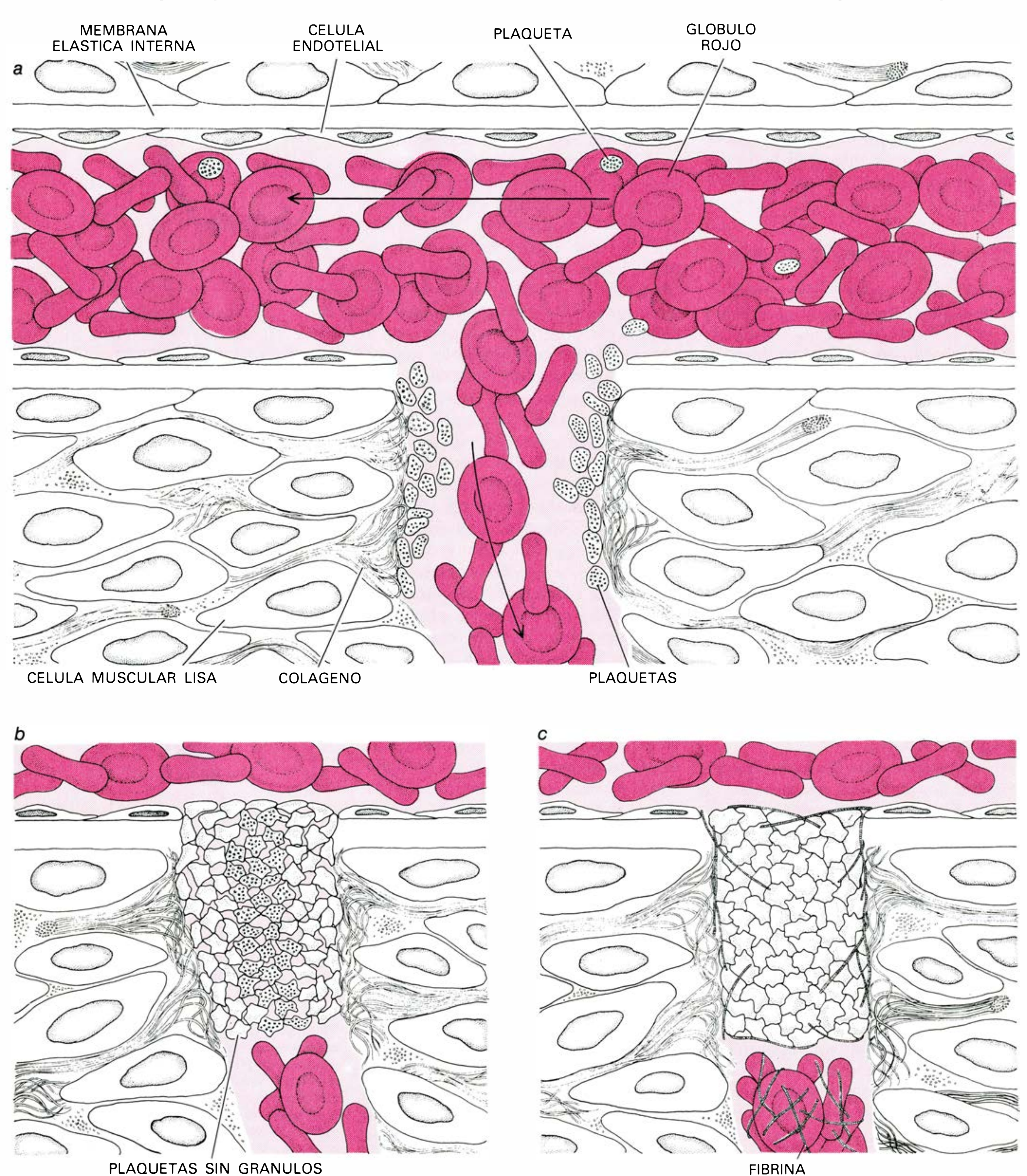
Una prueba desarrollada por Hans R. Baumgartner, Harvey J. Weiss y Thomas Tschopp, que trabajaban en



TRAZADOS de un agregómetro, que registra la agregación de las plaquetas en un plasma rico en plaquetas en respuesta a un agente agregante, en este caso el colágeno y a dos concentraciones distintas de ADP. El incremento inicial en la transmisión de luz se debe a la dilución por el agente añadido, la pequeña deflexión hacia abajo responde a un cambio en la forma de la plaqueta y el incremento subsiguiente se debe a la agregación. Con el colágeno se observa un período de latencia antes de que el ADP sea segregado por las plaquetas y produzca agregación; cuando el ADP se suministra directamente, el efecto es más rápido. El efecto de pequeñas concentraciones de ADP se neutraliza por enzimas que lo destruyen.

equipo en el laboratorio Hoffman-La Roche de Basilea y en el Hospital Roosevelt de Nueva York, ha servido para descifrar el mecanismo de interacción de las plaquetas con la pared del vaso en los pacientes que sufren la enfermedad de von Willebrand y en otras situaciones patológicas. Cuando sangre normal circula a través de anillos de arterias de conejo cuya capa endotelial ha sido eliminada para poner al descubierto la delgada capa subendotelial, que contiene lo que parece ser una forma especial de colágeno, el subendotelio queda cubierto de plaquetas en pocos minutos. La mayoría de las plaquetas pierden sus gránulos, pues segregan su contenido. A ciertos intervalos, a lo largo de la pared se forman pequeños agregados de plaquetas.

Las plaquetas se adhieren al colágeno y en varios aspectos la respuesta es semejante a la que tiene lugar en el



FUNCION DE LAS PLAQUETAS en la hemostasia, o detención del flujo sanguíneo en una herida. Tan pronto como la sangre empieza a fluir a través de un corte en la pared del vaso, las plaquetas se adhieren al colágeno de la pared (*a*). Las plaquetas son estimuladas y segregan el contenido de sus gránulos liberando sustancias como el ADP; otras plaquetas circulantes se adhieren a la primera capa constituyendo un tapón poco compacto que cierra provisionalmente la herida (*b*). Los cambios que experimentan las plaquetas y el contacto de la sangre con células dañadas transforman la proteína plasmática protrombina en el enzima trombina. La trombina, a su vez, transforma el fibrinógeno en hebras de fibrina, que estabilizan el tapón (*c*).

agregómetro. Por ejemplo, los agregados no se forman cuando la arteria es perfundida con sangre tratada con aspirina o de pacientes con tromboastenia. El sistema de perfusión de Baumgartner revela sin embargo la necesidad del factor de von Willebrand, lo que no resultaba evidente en los estudios realizados con el agregómetro. El factor debe estar presente para que las plaquetas se adhieran al tipo especial de colágeno subyacente a la capa endotelial. No se sabe con certeza si el factor de von Willebrand es necesario por la gran velocidad del flujo sanguíneo o por alguna peculiaridad especial del colágeno subepitelial. El factor está presente en las células endoteliales, en las plaquetas y también en el plasma sanguíneo, que parece ser la fuente principal en la hemostasia.

Otros datos en favor de la importancia del factor de von Willebrand proceden del estudio de otra alteración congénita, el síndrome de las plaquetas gigantes de Bernard-Soulier. En estos enfermos las plaquetas gigantes no se adhieren al subendotelio de los anillos arteriales perfundidos y no se agrupan en presencia del factor de von Willebrand; falta el receptor para dicho factor.

Patología de las plaquetas

Los errores de la naturaleza manifiestos en enfermedades como la tromboastenia y el síndrome de Bernard-Soulier proporcionan información sobre los mecanismos de la función plaquetaria normal. Por ejemplo, la falta de gránulos densos en las plaquetas de los pacientes con tromboastenia puede correlacionarse con niveles bajos de nucleótidos de adenina y serotonina en dichas plaquetas. Las alteraciones de la agregación en presencia de colágeno y el prolongado tiempo de sangría en estos pacientes demuestra la importancia de los gránulos densos en la hemostasia normal. En casos raros se ha observado una alteración denominada "síndrome plaquetario gris", en el que las plaquetas carecen de gránulos alfa. Se están realizando actualmente estudios sobre la función de tales plaquetas, lo que presumiblemente permitirá descifrar el papel del contenido de dichos gránulos.

En la tromboastenia y en el síndrome de Bernard-Soulier los defectos radican en glicoproteínas expuestas en la superficie de las plaquetas. Las proteínas de la membrana pueden separarse en distintas bandas por electroforesis y las glicoproteínas pueden identificarse utilizando colorantes para los hidratos de carbono. Las proteínas expuestas en la superficie de la membrana pueden identificarse haciéndolas reaccionar con compuestos marcados radiactivamente que se unen a la membrana pero que no pueden atravesarla. Al romper las plaquetas y someter las proteínas de la membrana a electroforesis, las bandas radiactivas identifican las proteínas de la superficie. Las plaquetas normales muestran tres bandas principales correspondientes a glicoproteínas de la superficie designadas GP I, GP II y GP III en orden de tamaño decreciente. Cada banda probablemente esté constituida por varias glicoproteínas; investigadores de distintos laboratorios tratan actualmente de caracterizarlas.

Las electroforesis de plaquetas de pacientes con tromboastenia muestran una menor cantidad de proteína en las bandas GP II y GP III. Las enfermedades congénitas suponen en general el déficit de un gen y de una proteína, por lo que sorprende que en esta enfermedad sean dos las proteínas deficitarias. Investigaciones recientes sugieren no obstante que la separación podría ser un artefacto y el déficit genético afectaría únicamente a una proteína. Quizá llegue a demostrarse que la proteína deficiente es el receptor del fibrinógeno, pues las plaquetas de los pacientes con tromboastenia ni se agregan ni se unen al fibrinógeno. En el síndrome de Bernard-Soulier falta parte de la banda GP I. Podría tratarse de la ausencia del receptor para el factor de von Willebrand, con el que dichas plaquetas no reaccionan.

Terapia

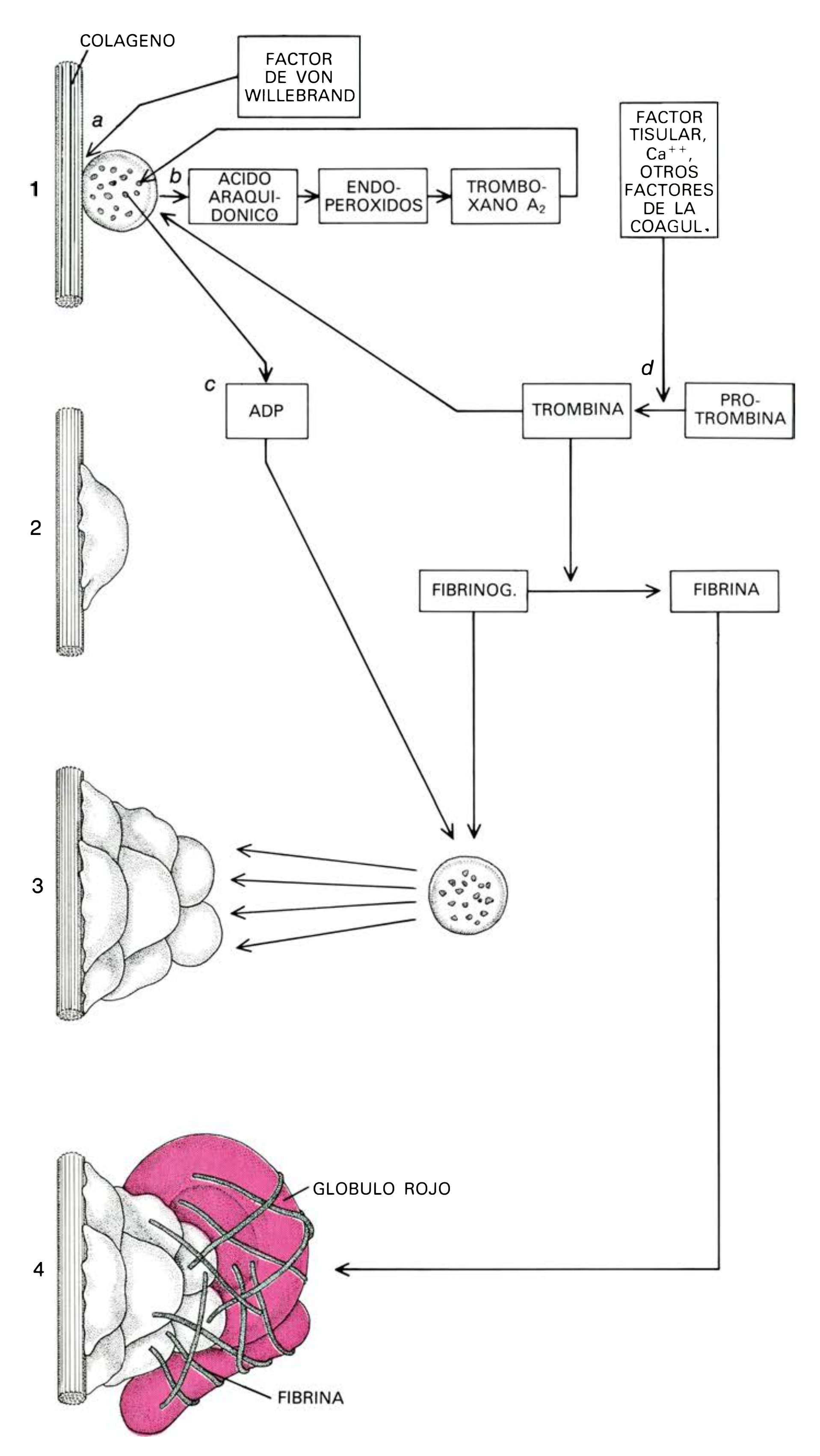
La transfusión desempeña un papel importante en el tratamiento de las enfermedades plaquetarias. La sangre de los donantes se fracciona habitualmente en distintas partes, cada una de las cuales tiene un valor terapéutico especial. El plasma sirve para incrementar el volumen de sangre, los glóbulos rojos para tratar la anemia, distintas proteínas plasmáticas son utilizadas para el tratamiento de enfermedades por déficit específicos tales como la hemofilia o la enfermedad de von Willebrand. Una dificultad particular de las plaquetas radica en que, a diferencia del plasma y de los glóbulos rojos, no pueden almacenarse durante semanas y deben transfundirse pocos días después de la donación de la sangre. Además, un paciente que requiera muchas transfusiones desarrollará anticuerpos frente a las plaquetas que no sean de su mismo tipo. Desafortunadamente, con las plaquetas, al igual que con los tejidos transplantados, resulta difícil evitar las incompatibilidades. En el caso de los glóbulos rojos la mayoría de las veces basta tipificarlos como *Rh* positivos o *Rh* negativos y encuadrarlos en los cuatro grupos *A, B, O* o *AB*. Las plaquetas difieren marcadamente de una persona a otra, siendo las plaquetas de un familiar próximo al paciente las más compatibles.

Las transfusiones de plaquetas son útiles cuando el recuento plaquetario es bajo, tal como ocurre en la trombocitopenia causada por un defecto de producción y cuando las plaquetas son defectuosas, como en el caso del síndrome de Bernard-Soulier. Producen poca mejoría en la púrpura trombocitopénica idiopática, ya que los anticuerpos del paciente destruyen las plaquetas extrañas tanto como las propias. Afortunadamente, algunos tratamientos inmunosupresores, tales como la administración de esteroides o la extirpación del bazo, resultan efectivos, recuperándose habitualmente el paciente.

Trombosis

Puesto que las plaquetas pueden adherirse entre sí y a las paredes de los vasos sanguíneos, y debido a que promueven la coagulación, desempeñan un papel central en la formación de los trombos, o coágulos intravasculares, que interrumpen el flujo de sangre. Los trombos se forman a menudo en las venas, especialmente en personas de edad avanzada que permanecen en cama después de una intervención quirúrgica. Estos trombos pueden ocluir una vena, pero resultan especialmente peligrosos por su tendencia a desprenderse para dar lugar a émbolos que son arrastrados por la corriente sanguínea a través del corazón y de la arteria pulmonar para alojarse en una pequeña arteria del pulmón, a la que obstruyen. Los trombos venosos están compuestos por capas alternantes de masas plaquetarias y de hebras de fibrina con glóbulos rojos atrapados. Los trombos se forman también en las arterias; estos trombos contienen una mayor proporción de plaquetas y tienen un aspecto más grisáceo que rojo. La oclusión de una arteria por un trombo es peor tolerada por el organismo que la oclusión de una vena, sobre todo cuando la arteria irriga una región vital como el cerebro o el músculo cardíaco.

La causa más común de trombos arteriales es la enfermedad conocida como arteriosclerosis o aterosclerosis,



INTERACCIONES COMPLEJAS de las plaquetas, constituyentes plaquetarios y otras sustancias implicadas en la formación del tapón hemostático y de los trombos. El mecanismo de muchas de estas interacciones no es aún conocido. El contacto de una plaqueta con el colágeno (*1*) en presencia del factor de von Willebrand (*a*) inicia una vía (*b*) que estimula la secreción de ADP a partir de los gránulos densos (*c*). La plaqueta que se adhiere cambia de forma, se extiende a lo largo del colágeno y pierde los gránulos (*2*). Al mismo tiempo, una serie de pasos que implican al factor tisular, a los iones calcio y otros factores de la coagulación transforman la protrombina del plasma en trombina (*d*), que se forma así mismo en la superficie de las plaquetas. La trombina estimula también la secreción y transforma el fibrinógeno (del plasma y de las plaquetas) en fibrina. Bajo la influencia del colágeno, ADP y trombina se produce la agregación plaquetaria (*3*). Las hebras de fibrina estabilizan el tapón. El proceso puede detenerse en este momento o proseguir dando lugar a un trombo de mayor tamaño con glóbulos rojos atrapados (*4*).

que se caracteriza por un marcado engrosamiento de la capa más interna de la pared arterial: la íntima. Esta capa está constituida normalmente por un solo estrato monocelular de células endoteliales y un delgado subendotelio limitado en la parte externa por la membrana elástica interna. En la arterioesclerosis, la íntima resulta invadida por células musculares lisas y experimenta un engrosamiento. A medida que progresa la enfermedad, se producen grandes depósitos de lípidos en la parte interna de la pared del vaso, con lo que resulta aún más reducida la luz arterial. La arteriosclerosis es la causa más común de los ataques cardíacos. En la pared de las arterias coronarias estrechadas se forman grandes trombos que dificultan o bloquean el flujo de sangre; pequeños trombos pueden soltarse y circular como émbolos que se alojan en arterias más pequeñas. Ambos tipos de trombos pueden bloquear la circulación coronaria y conducir al infarto miocárdico, la muerte de parte del músculo cardíaco. La apoplejía, por otra parte, está también a menudo relacionada con la arteriosclerosis de las arterias carótidas. Un trombo o un émbolo reduce el flujo sanguíneo en determinadas partes del cerebro, a veces de forma abrupta.

En los últimos 15 años, varios grupos de investigadores, entre ellos el de J. F. Mustard, del Centro Médico de la Universidad McMaster, y el de Russell Ross, de la Facultad de Medicina de la Universidad de Washington, han demostrado que las plaquetas están implicadas en el desarrollo de la arteriosclerosis, como lo están en la formación de los trombos. Se considera que el primer fenómeno consistiría en una lesión de las células endoteliales. La lesión podría originarse a raíz del impacto y la turbulencia del flujo sanguíneo, particularmente en los puntos donde las arterias se ramifican. Otra posibilidad consiste en que la lesión tenga origen bioquímico. El monóxido de carbono y otras sustancias existentes en el aire contaminado de las ciudades y en el humo de los cigarrillos, así como lípidos de la dieta y hormonas (como la adrenalina) producidas en el organismo en situaciones de stress, contribuirían globalmente al desarrollo de la arteriosclerosis, y lo harían alterando las células endoteliales, que al desprenderse se dejarían al descubierto el subendotelio.

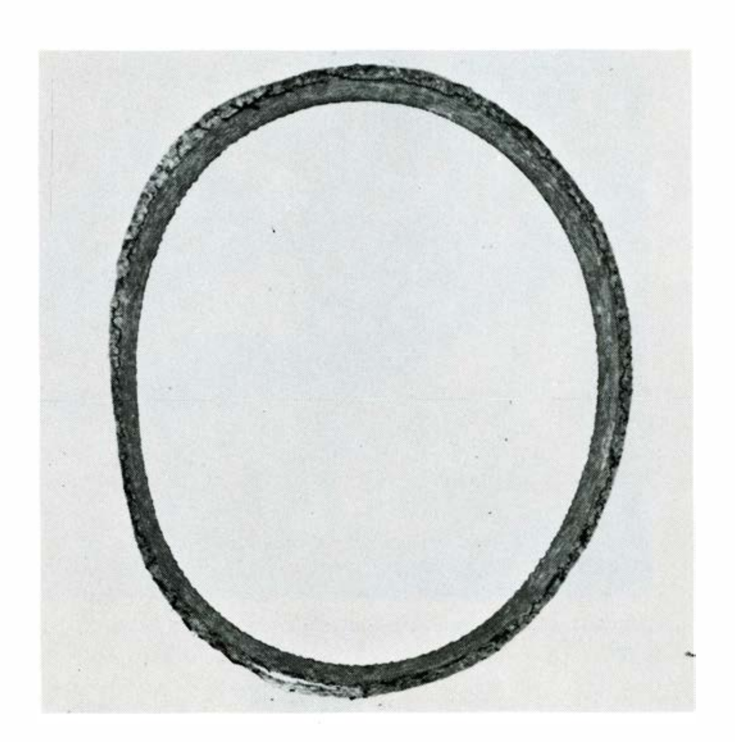
Baumgartner ha desarrollado un sistema, a partir de la aorta del conejo, para estudiar la hemostasia y la génesis de la arteriosclerosis. Las células endoteliales pueden desprenderse del vaso

mediante la introducción de un balón, unido al final de un catéter, en la aorta de un conejo anestesiado. Después de hinchar el balón, su desplazamiento a través de la aorta provoca el desprendimiento del endotelio; el conejo se recupera rápidamente. En pocos minutos, la aorta queda cubierta de plaquetas. Al cabo de pocos días, la íntima empieza a engrosarse y las células musculares lisas migran a través de soluciones de continuidad de la membrana elástica interna y experimentan divisiones, con lo que se estrecha la luz del vaso. Las células de la musculatura lisa, al igual que las células endoteliales, no provocan la adhesión de las plaquetas ni el desarrollo de trombos. Theodore H. Spaet y Michael B. Stemerman, del Hospital Montefiore de Nueva York, han mostrado, sin embargo, que un traumatismo repetido con el balón conduce a un mayor engrosamiento. Quizás agresiones permanentes sobre el endotelio vascular provoquen el mismo efecto en los seres humanos.

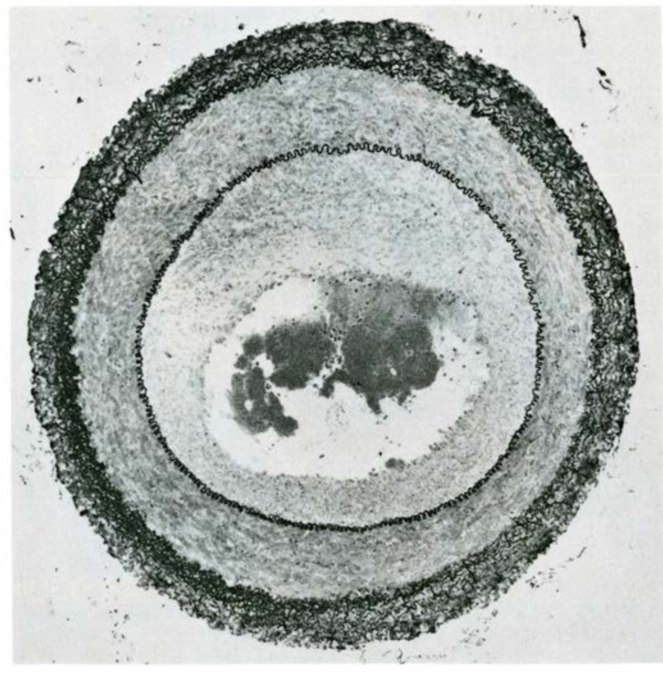
Plaquetas y arteriosclerosis

Un estímulo esencial para la migración y división de las células musculares lisas podría ser el factor de crecimiento plaquetario, una de las sustancias segregadas a partir de los gránulos alfa. Dos observaciones en animales de experimentación, aún no confirmadas, sugieren tal función. Una reducción drástica del número de plaquetas de los conejos antes de aplicar el balón en la aorta evitaba el engrosamiento subsiguiente de la pared arterial. La arteriosclerosis experimental no se desarrollaba en cerdos que padecían una forma grave de la enfermedad de von Willebrand, a diferencia de los cerdos normales, cuando recibían una dieta de alto contenido en colesterol. De ello se deducía que en ausencia del factor de von Willebrand, las plaquetas no se adherían a la pared arterial dañada por el colesterol.

Se dispone de abundantes datos experimentales que prueban que las pla-



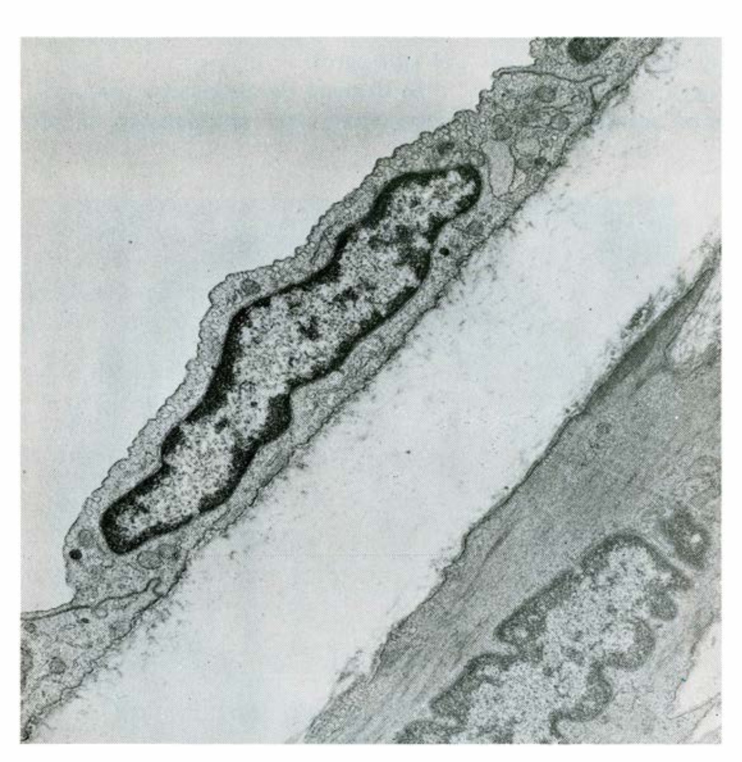


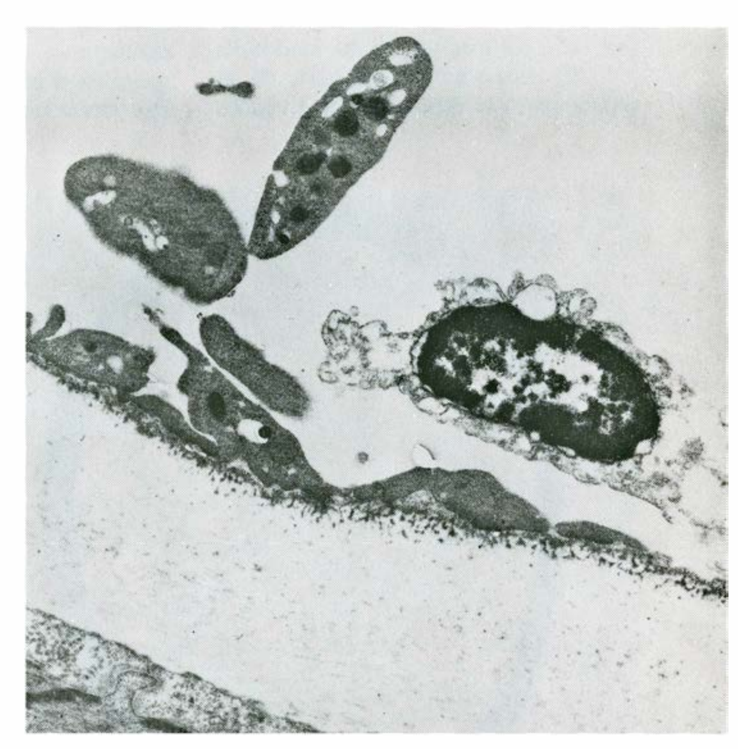


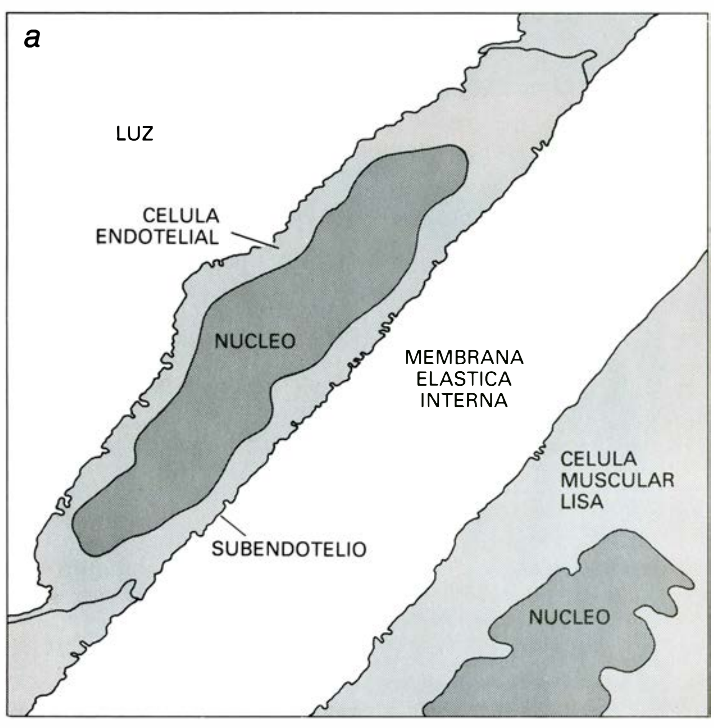
ARTERIOSCLEROSIS Y TROMBOSIS en un modelo experimental en la aorta de conejo diseñado por Baumgartner. La aorta normal *(arriba a la izquierda)* está constituida fundamentalmente por tejido conectivo y células de musculatura lisa, limitadas en la parte interna por la membrana elástica interna, que apenas se aprecia como una línea oscura, ondulada en algunos puntos. La capa más interna de la membrana, la íntima, consta de un estrato monocelular de células endoteliales y un delgado subendotelio rico en colágeno; la íntima no puede observarse a estos aumentos. Después de eliminar la capa endotelial, se producen una serie de fenómenos, iniciados con un depósito de plaquetas, que conducen al engrosamiento de la íntima *(arriba, a la derecha)*, mientras células de la musculatura lisa atraviesan la membrana elástica interna *(línea ondulada oscura)* y proliferan. Los trombos plaquetarios pueden ocluir la luz estrechada del vaso.

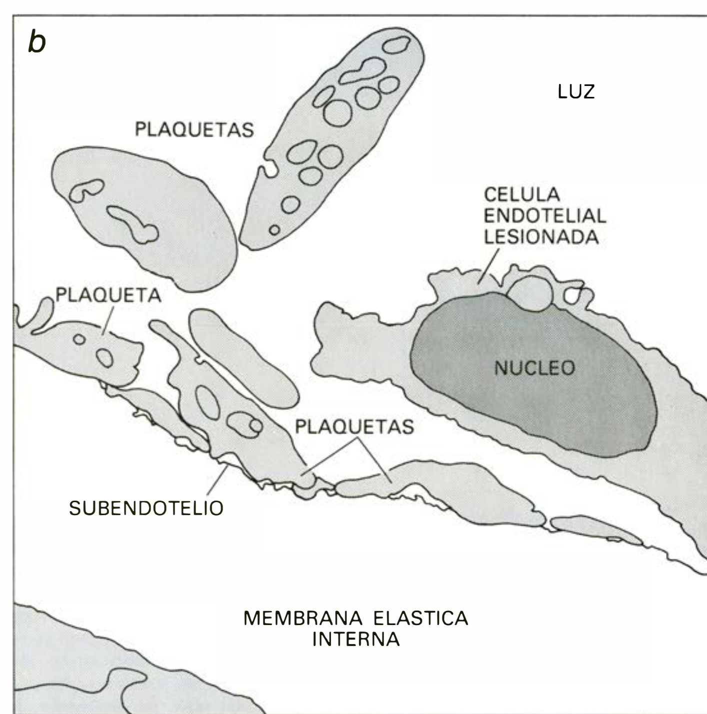
quetas se comportan de forma anormal en personas con arteriosclerosis grave. Las plaquetas de algunos de estos enfermos tienen una vida media más corta. Para determinar la vida media de las plaquetas, se incuba una muestra de sangre con cromo radiactivo que permite marcarlas sin que resulte afectada su función. Las plaquetas marcadas se inyectan. En sujetos normales, alrededor de una décima parte de las plaquetas marcadas (y presumiblemente de las no marcadas) desaparece diariamente de la circulación, indicando que la supervivencia normal es de unos 10 días. En muchas personas que padecen arteriosclerosis grave, la supervivencia de las plaquetas es sólo de seis o siete días. Al parecer, las plaquetas o se adhieren a la pared arterial lesionada, o se alteran por contacto con ella, siendo eliminadas de la circulación en otro punto.

Otros datos sobre la afectación plaquetaria se obtienen por determinaciones de productos plaquetarios en el plasma. Además del factor de crecimiento, al menos otras dos sustancias, el factor plaquetario 4 y la beta-tromboglobulina, características de los gránulos plaquetarios, se encuentran sólo en pequeñísimas cantidades en el plasma separado cuidadosamente de sangre normal tratada de forma que resulte inhibida la secreción plaquetaria. El plasma de enfermos con trombosis o arteriosclerosis grave a menudo contiene cantidades mucho mayores de ambas sustancias, lo que indica que sus







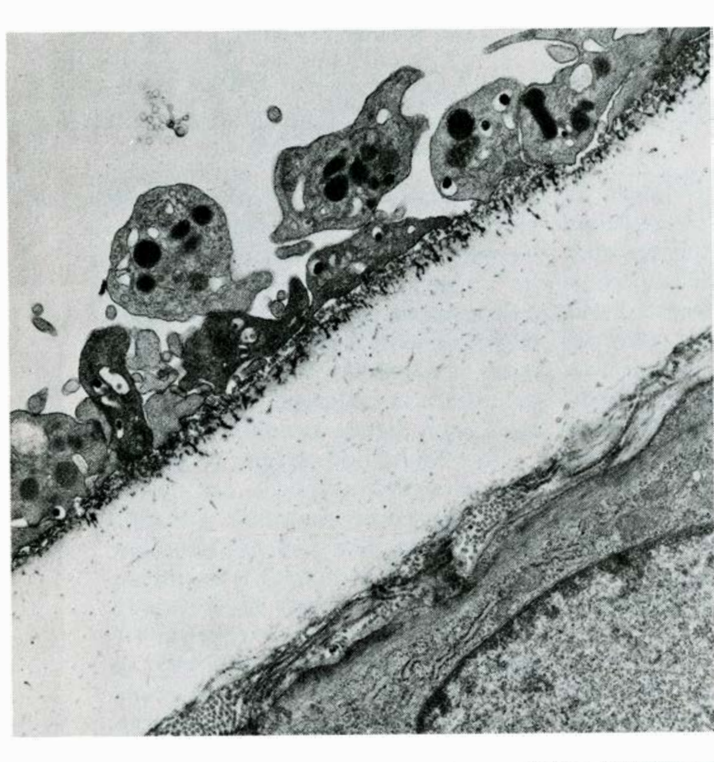


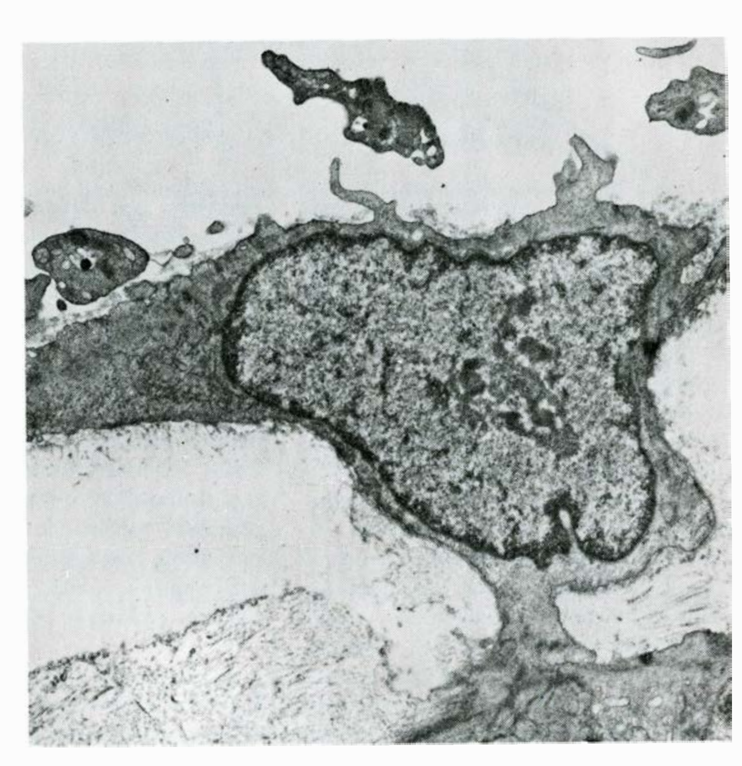
CAMBIOS EN LA INTIMA de la aorta del conejo a raíz de lesiones producidas en las células endoteliales. Las micrografías electrónicas fueron realizadas por Baumgartner y las distintas estructuras están identificadas en los dibujos, debajo de cada micrografía; la ampliación en cada caso es de unos 10.000 diámetros. En la pared arterial intacta (*a*) pueden observarse las células endoteliales que delimitan la luz del vaso, la membrana elástica interna

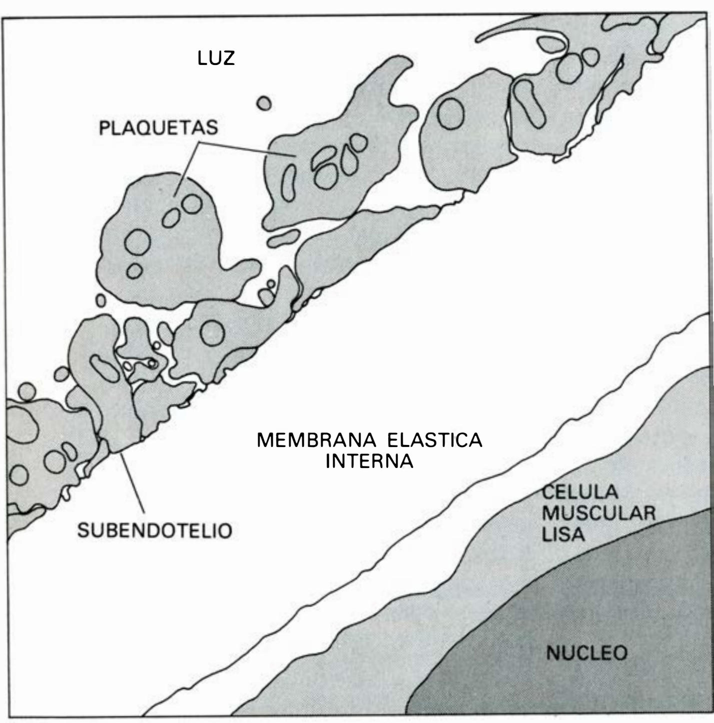
plaquetas se están rompiendo continuamente o bien segregan el contenido de sus gránulos a la circulación.

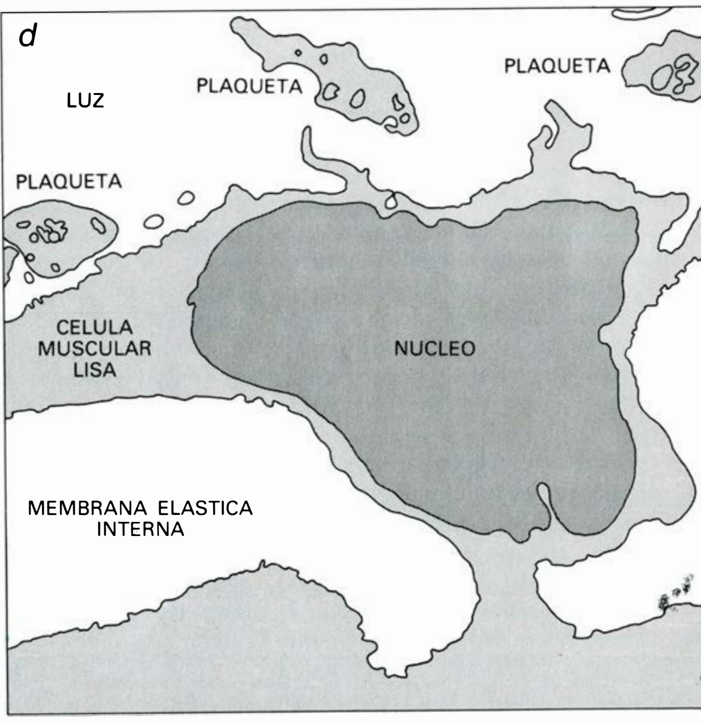
En muchos laboratorios se está actualmente investigando la función de las plaquetas en la arteriosclerosis, en la trombosis y en enfermedades como la diabetes, en la que los dos procesos anteriores constituyen complicaciones frecuentes. Estos estudios implican no sólo la determinación del tiempo de supervivencia de las plaquetas o la determinación de los niveles plasmáticos de productos plaquetarios, sino también el estudio de los efectos de determinadas drogas en dichas variables y en la incidencia de ataques cardíacos y otras secuelas de la arterioesclerosis. Algunas drogas, como el dipiridamol y la sulfinpirazona, prolongan al parecer la vida media de las plaquetas que en otras circunstancias tendrían una supervivencia corta. El descubrimiento de este efecto fue fortuito y el mecanismo de acción es desconocido. Otras drogas, como la aspirina, cuyo modo de acción se conoce mejor, no son tan eficaces a la hora de prolongar la supervivencia de las plaquetas.

La investigación fisiológica y bioquímica, los estudios clínicos aplicados, pueden contribuir a una mejor comprensión del papel que las plaquetas desempeñan en la arteriosclerosis, la primera causa de muerte en las sociedades avanzadas, así como en otras enfermedades y alteraciones congénitas.









y parte de una célula muscular lisa por debajo de ella. Cuando una célula endotelial dañada se desprende (*b*), las plaquetas se adhieren al subendotelio y entre sí hasta formar una capa que cubre el subendotelio (*c*). Entonces, una célula muscular lisa migra a través de una solución de continuidad de la membrana elástica interna (*d*); con el tiempo, las células musculares lisas forman una gruesa pseudoíntima que puede provocar un trombo.

Teorías gauge de las fuerzas entre partículas elementales

Las fuerzas básicas de la naturaleza se describen ahora mediante teorías de esta clase. Las propiedades de las fuerzas se deducen de simetrías o regularidades manifiestas en las leyes de la física

Gerard 't Hooft

Para poder entender cómo está constituido el mundo necesitamos una teoría que explique las interacciones entre las partículas elementales de la materia. O lo que es lo mismo: se requiere una teoría de las fuerzas básicas de la naturaleza. Se han identificado cuatro fuerzas de esta clase; y, hasta hace poco, cada una de ellas precisaba una teoría distinta. La fuerza de la gravitación y la del electromagnetismo tienen un alcance ilimitado: razón principal de que las conozca todo el mundo. Pueden sentirse directamente como causantes de atracciones o repulsiones. Las restantes fuerzas, denominadas fuerza débil y fuerza fuerte, escapan a la percepción directa porque su influencia se extiende sólo a distancias cortas, no mayores que el radio del núcleo atómico. La fuerza fuerte mantiene unidos los protones y los neutrones en el núcleo y, en otro contexto, liga las partículas llamadas quarks que se cree que son los constituyentes de protones y neutrones. La fuerza débil es responsable en gran parte de la desintegración de ciertas partículas.

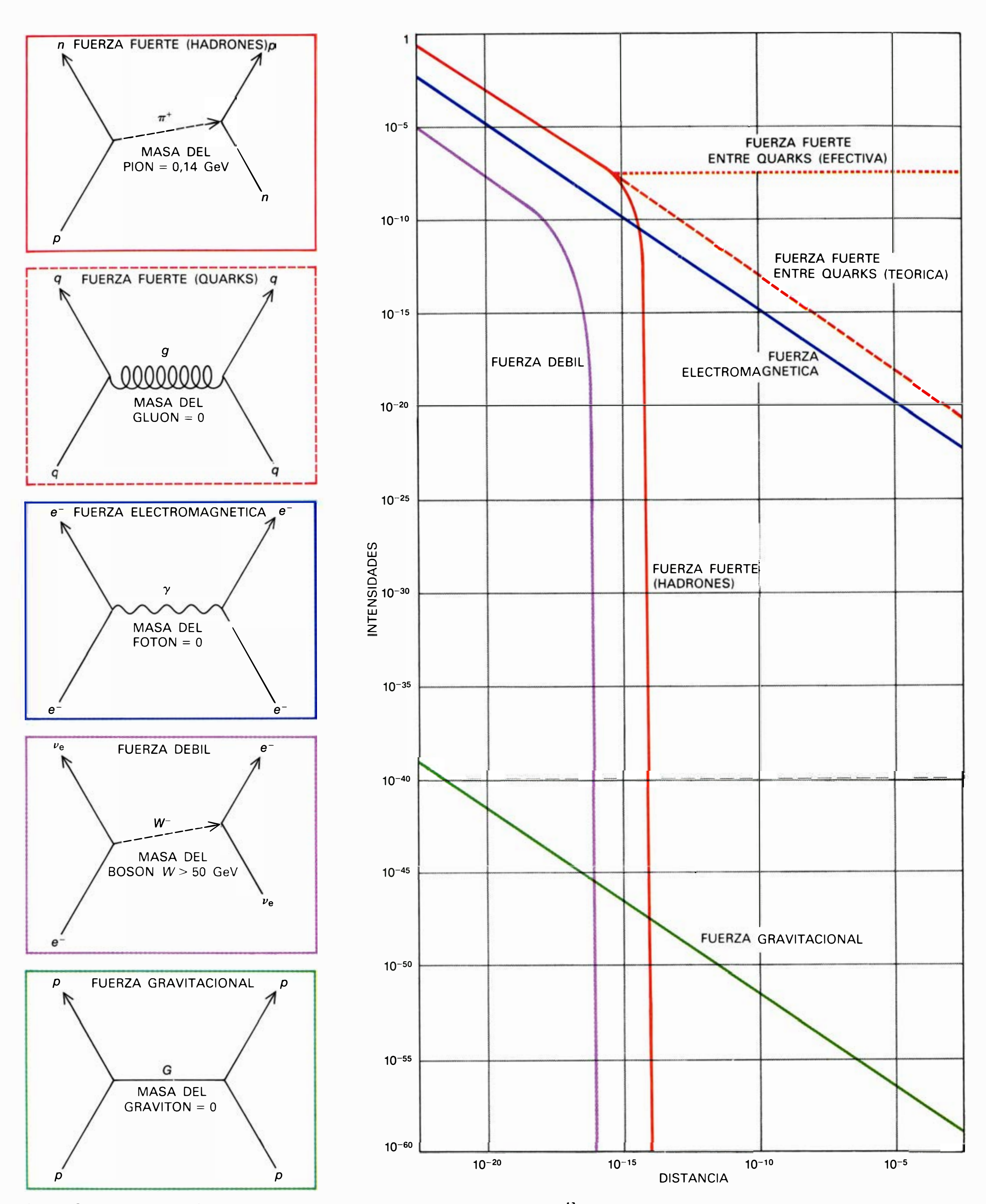
Una ambición de los físicos, que se ha ido afirmando con el tiempo, es llegar a formular una teoría central única que incorporara todas las fuerzas conocidas. Una teoría que revelara profundas conexiones entre los distintos tipos de fuerzas a la vez que diera cuenta de su diversidad aparente. Aunque no se ha logrado tal unificación, en los últimos años se han realizado, qué duda cabe, algunos progresos. Podemos ya interpretar la fuerza débil y el electromagnetismo en el contexto de una teoría única. No se trata de que ambas fuerzas pierdan su identidad característica, sino de que la teoría las implica matemáticamente. Y lo que a la larga puede resultar más importante es que ahora las cuatro fuerzas vienen descritas por teorías que tienen la misma forma general. Los físicos deben seguir buscando la llave que les abra todas las cerraduras conocidas; pero ya saben, al menos, que todas las llaves necesarias pueden cortarse a partir del mismo molde. Las teorías pertenecientes a esta clase privilegiada se denominan formalmente teorías gauge no abelianas con simetría local. Qué significa tan rimbombante título constituye el principal motivo del artículo. Baste, por ahora, con adelantar que las teorías relacionan las propiedades de las fuerzas con simetrías de la naturaleza.

Desde Galileo y Newton, las simetrías y las simetrías aparentes de las leyes de la naturaleza han intervenido en la construcción de las teorías físicas. Las simetrías más familiares son las espaciales o geométricas. En un copo de nieve, por ejemplo, la presencia de cierta simetría puede detectarse sin dificultad I a simetría podría definirse como una invariancia de la figura observada al aplicarle una determinada transformación. En el caso del copo de nieve, la transformación es una rotación de 60 grados, o un sexto de círculo. Si marcamos la posición inicial y giramos luego el copo de nieve 60 grados (o cualquier múltiplo entero de 60 grados), no se percibirá ningún cambio. El copo de nieve es invariante con respecto a las rotaciones de 60 grados. De acuerdo con el mismo principio, un cuadrado es invariante bajo rotaciones de 90 grados y de un círculo se dice que tiene una simetría continua porque una rotación de cualquier ángulo lo deja invariante.

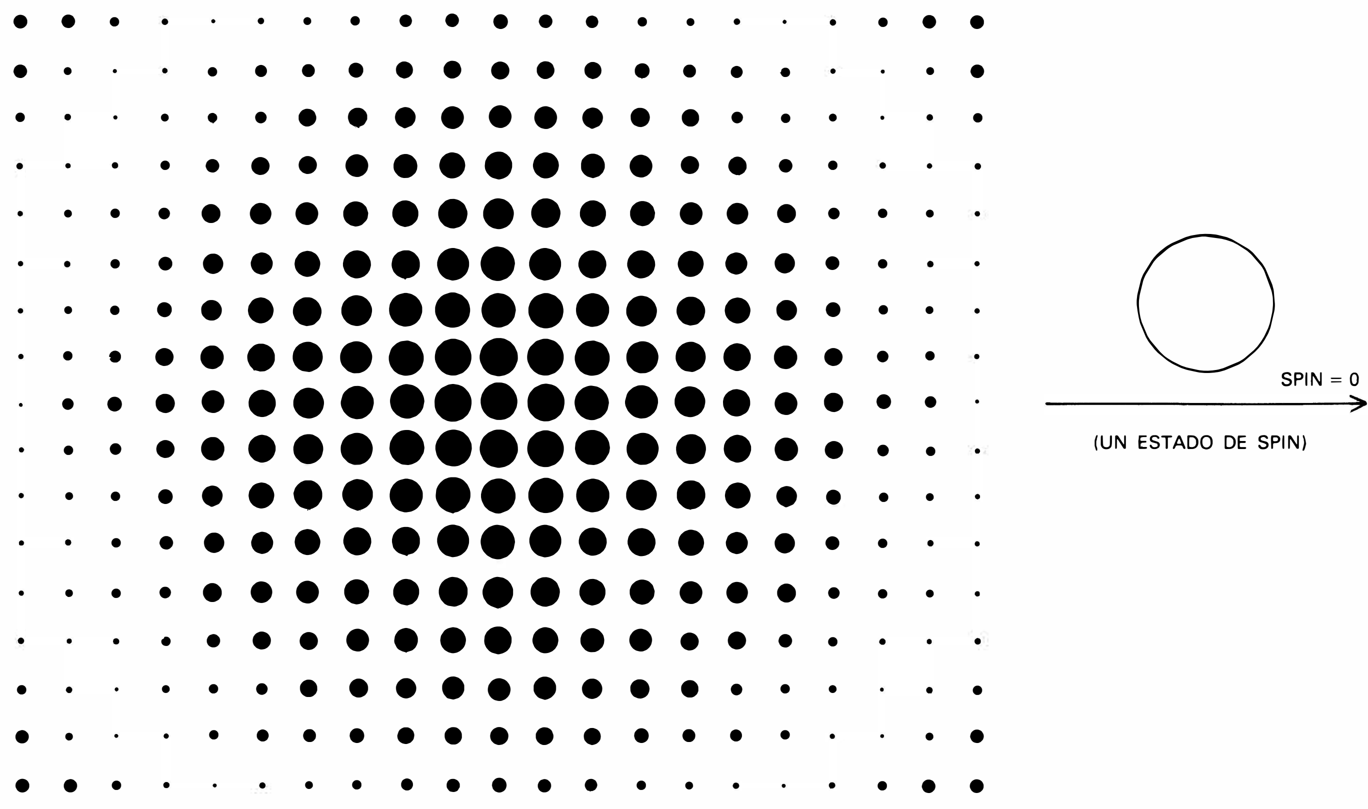
Si bien el concepto de simetría tiene su origen en la geometría, es lo bastante general para abarcar invariancias con respecto a transformaciones de otras clases. Un ejemplo de simetría no geométrica es la simetría de carga del electromagnetismo. Supóngase que se ha colocado cierto número de partículas cargadas eléctricamente en una configuración dada y que se han medido todas las fuerzas que actúan entre pares de partículas. Si, a continuación, se invierte la polaridad de las cargas, las fuerzas siguen siendo las mismas.

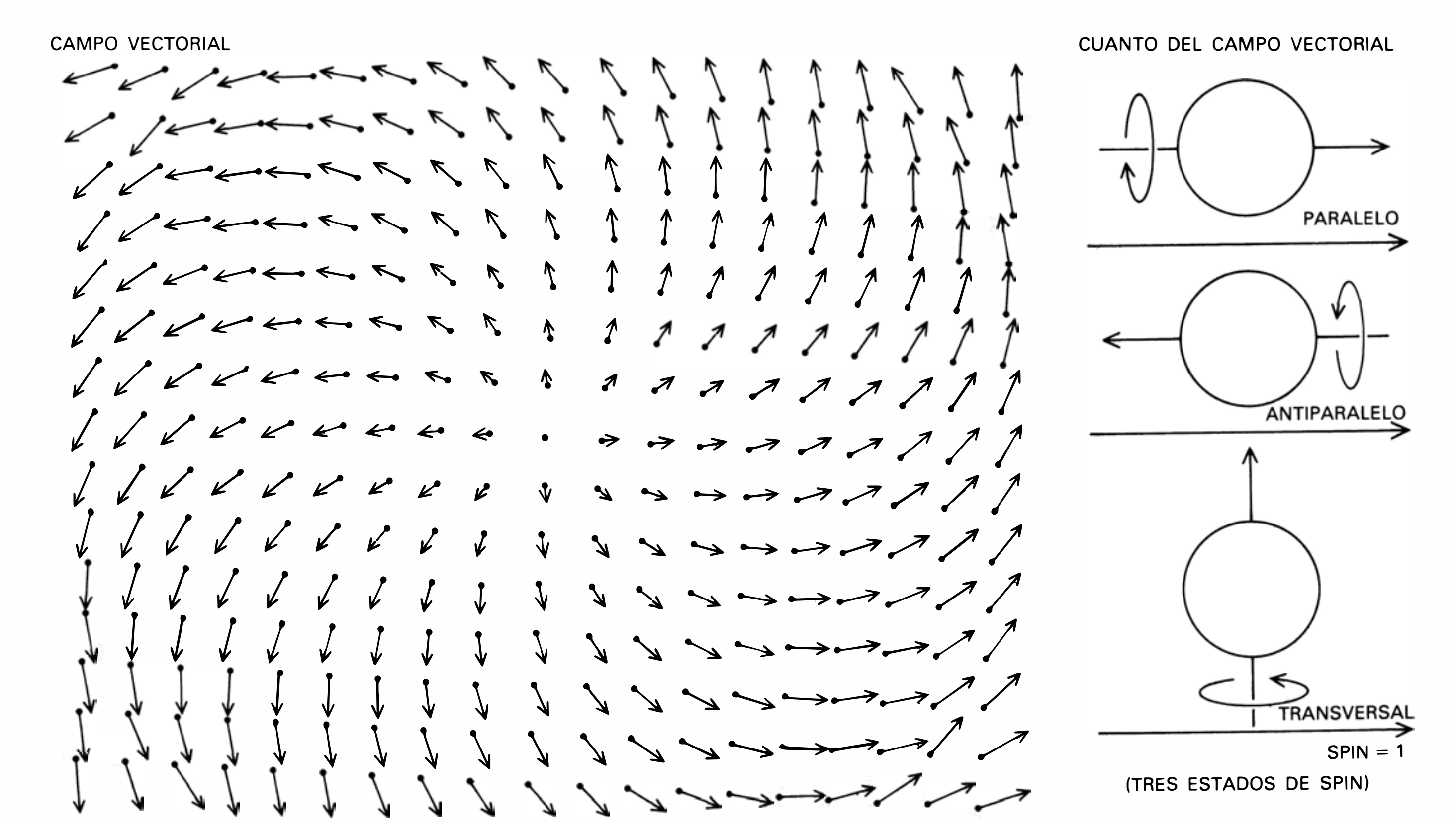
Otra simetría de carácter no geométrico es el spin isotópico, una propiedad de protones y neutrones y de otras muchas partículas llamadas hadrones, los únicos que responden a las interacciones fuertes. La base de la simetría yace en el hecho de que el protón y el neutrón son dos partículas muy semejantes. Difieren en masa en un uno por mil aproximadamente, y, si exceptuamos su carga eléctrica, son idénticos en todas las demás propiedades. Parece, pues, que protones y neutrones podrían intercambiarse sin quedar prácticamente afectadas las interacciones fuertes. Si las fuerzas electromagnéticas (que dependen de la carga eléctrica) pudieran eliminarse de alguna forma, la simetría de spin isotópico sería exacta; en realidad sólo es aproximada.

Aunque el protón y el neutrón parecen ser partículas distintas, y es difícil imaginar un estado de la materia intermedio entre ellos, resulta que la simetría con respecto al spin isotópico es una simetría continua, más parecida a la simetría de la esfera que a la del copo de nieve. Daré una explicación simplificada de por qué ello es así. Imaginemos que en el interior de cada partícula hay un par de flechas cruzadas, una que represente la componente protón de la partícula y la otra que simbolice la componente neutrón. Si la flecha protónica está orientada hacia arriba (no importa qué dirección definamos hacia arriba), la partícula es un protón; si la



CUATRO FUERZAS BASICAS median todas las interacciones conocidas entre las partículas materiales. Las fuerzas difieren mucho en intensidad y en alcance efectivo, pero vienen descritas por teorías con la misma estructura matemática, las llamadas teorías gauge locales. Se dice que el electromagnetismo y la gravitación tienen alcance infinito, aunque su influjo decrece como el cuadrado de la distancia entre dos partículas que interaccionan. La fuerza débil está confinada a un alcance extraordinariamente corto, de unos 10^{-15} centímetros. Las propiedades de las interacciones fuertes son algo más complicadas. Mientras la fuerza fuerte se observa actuando entre hadrones, tales como el protón y el neutrón (*línea sólida coloreada*), tiene un alcance de unos 10^{-13} centímetros. La fuerza fuerte también liga entre sí las partículas denominadas quarks que componen los hadrones; en cuyo contexto se espera que sigan una ley del inverso del cuadrado de la distancia (*línea coloreada a trazos*). El comportamiento real es, al parecer, más raro: la fuerza permanece constante con independencia de la distancia (*línea coloreada de puntos*). En las teorías cuánticas de campos (*diagramas de la izquierda*), la fuerza entre dos partículas se manifiesta por el intercambio de una tercera partícula, que se llama partícula virtual. El alcance de la fuerza viene determinado por la masa de la partícula virtual intercambiada. Partículas virtuales sin masa, como el fotón y el gravitón, dan origen a fuerzas que tienen un alcance infinito.





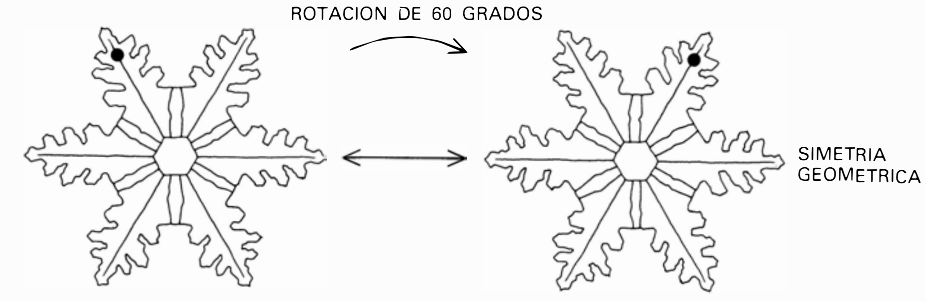
EL CONCEPTO DE CAMPO, una cantidad definida en cada punto de una cierta región del espacio y del tiempo, es importante en la construcción de las teorías gauge. Un campo escalar tiene una sola magnitud en cada punto. En este ejemplo la magnitud viene dada por el área de los puntos. Un campo vectorial posee magnitud y dirección; se ilustra dibujando una flecha en cada punto. Un campo escalar representa cantidades como la temperatura o la densidad de un fluido, en tanto que un campo vectorial representa su velocidad. En teorías cuánticas de campos, el influjo de un campo puede incorporarse mediante una partícula virtual. El número de componentes de un campo viene reflejado en el número de orientaciones distintas de la partícula, que a su vez depende de su momento angular de spin. Un campo escalar tiene una sola componente (su valor se puede dar con un solo número) y se representa por una partícula de spin cero con un único estado de spin. Un campo vectorial en el espacio tridimensional tiene tres componentes.

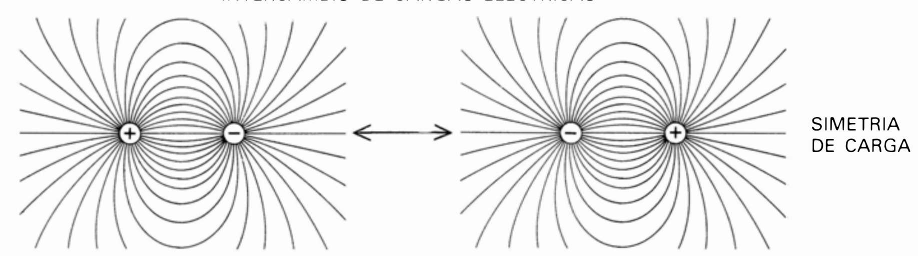
flecha neutrónica está hacia arriba, la partícula es un neutrón. Las posiciones intermedias corresponden a superposiciones mecánico-cuánticas de los dos estados; la partícula parece, entonces, unas veces un protón y otras un neutrón. La transformación de simetría asociada con el spin isotópico gira los indicadores internos de todos los protones y neutrones en cualquier lugar del universo con la misma cantidad y al mismo tiempo. Si la rotación es exactamente de 90 grados, cada protón se transforma en un neutrón y cada neutrón se transforma en un protón. La simetría bajo spin isotópico, hasta el punto en que sea exacta, establece que no se pueden detectar los efectos de esta transformación.

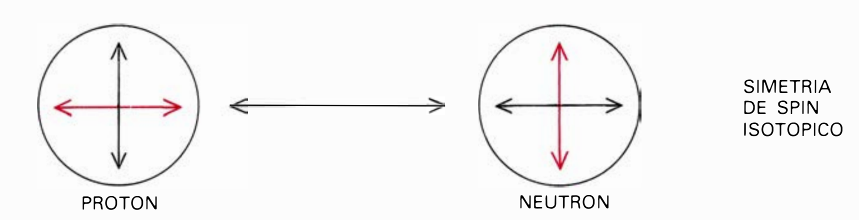
Todas las simetrías examinadas hasta aquí pueden caracterizarse como simetrías globales; en este contexto, la palabra global significa "que sucede por doquier y al mismo tiempo". En la descripción de la simetría de spin isotópico quedó explicitada esa restricción: la rotación interna que transforma protones en neutrones y neutrones en protones debe realizarse en cualquier punto del universo y en el mismo instante de tiempo. Además de las simetrías globales, que están casi siempre presentes en una teoría física, podemos obtener simetrías "locales"; en éstas, la convención puede decidirse, de forma independiente, en cada punto del espacio y en cada instante de tiempo. Aunque el término "local" sugiera, quizás, algo de alcance más modesto que una simetría global, la verdad es que la simetría local establece una condición mucho más exigente para la construcción de una teoría. Para la simetría global, ciertas leyes de la física permanecen invariantes cuando se aplica la misma transformación en el mismo instante y en todo lugar. Para que se observe una simetría local, la ley física debe mantener su validez aun cuando tenga lugar una transformación distinta en cada punto del espacio y del tiempo.

Se pueden construir teorías gauge con simetría global o local (o con ambas); las teorías con simetría local, sin embargo, encierran hoy el mayor interés. Para que una teoría resulte invariante con respecto a una transformación local debemos añadir algo nuevo: una fuerza. Pero antes de mostrar cómo ocurre esto, conviene exponer con cierto detalle el papel que ocupan las fuerzas en la descripción de las teorías modernas sobre las interacciones entre partículas elementales.

Los ingredientes básicos de una teoría de partículas son hoy, además de las partículas y las fuerzas, los campos. Se llama campo a una cantidad definida en todos los puntos de una región del espacio y el tiempo. La cantidad puede ser, por ejemplo, la temperatura y la región, la superficie de una sartén. El campo consta entonces de los valores de la temperatura para cada punto de la superficie.







SIMETRIAS DE LA NATURALEZA determinantes de las propiedades de las fuerzas en las teorías gauge. Reconocemos la simetría de un copo de nieve cuando observamos que la figura no cambia al dar un giro de 60 grados; se dice entonces que el copo de nieve es invariante bajo tales rotaciones. En física se introducen simetrías no geométricas. La simetría de carga, por ejemplo, es la invariancia de las fuerzas que actúan entre un conjunto de partículas cargadas cuando se invierten las polaridades de todas las cargas. La simetría de spin isotópico se basa en la observación de que apenas se registrarían cambios en las interacciones fuertes de la materia si se intercambiaran todos los protones y neutrones. Por tanto, el protón y el neutrón son sólo estados alternativos de una única partícula, el nucleón, y las transiciones entre los estados pueden realizarse (o imaginarse) ajustando la orientación de un indicador en un espacio interno. Esta clase de simetrías, donde la transformación consiste en una rotación interna o en un desplazamiento de fase, se denomina simetrías gauge. Desde Galileo y Newton, las simetrías y las simetrías aparentes de las leyes de la naturaleza han intervenido en la construcción de las teorías físicas.

Se dice que la temperatura es una cantidad escalar porque puede representarse por una posición a lo largo de una línea o escala. El campo de temperaturas correspondiente es un campo escalar, en el que cada punto viene asociado con un solo número o magnitud. Hay otras clases de campos, siendo los más importantes para nuestros propósitos los campos vectoriales, donde a cada punto del espacio se le hace corresponder un vector o flecha. Un vector tiene un tamaño, que se representa mediante la longitud de la flecha, y una dirección, que en espacios tridimensionales puede especificarse mediante dos ángulos; por tanto, hay que dar tres números para indicar el valor de un vector. Ejemplo de un campo vectorial es el campo de velocidades en un fluido; en cada punto del volumen ocupado por el fluido se dibuja una flecha para designar la velocidad y la dirección del flujo.

En la física de los objetos cargados eléctricamente, el campo constituye un modo adecuado de expresar cómo la fuerza del electromagnetismo se traslada de un lugar a otro. Se supone que de

todas las partículas cargadas emana un campo electromagnético; cada partícula interacciona luego con la suma de todos los campos en vez de hacerlo directamente con las otras partículas.

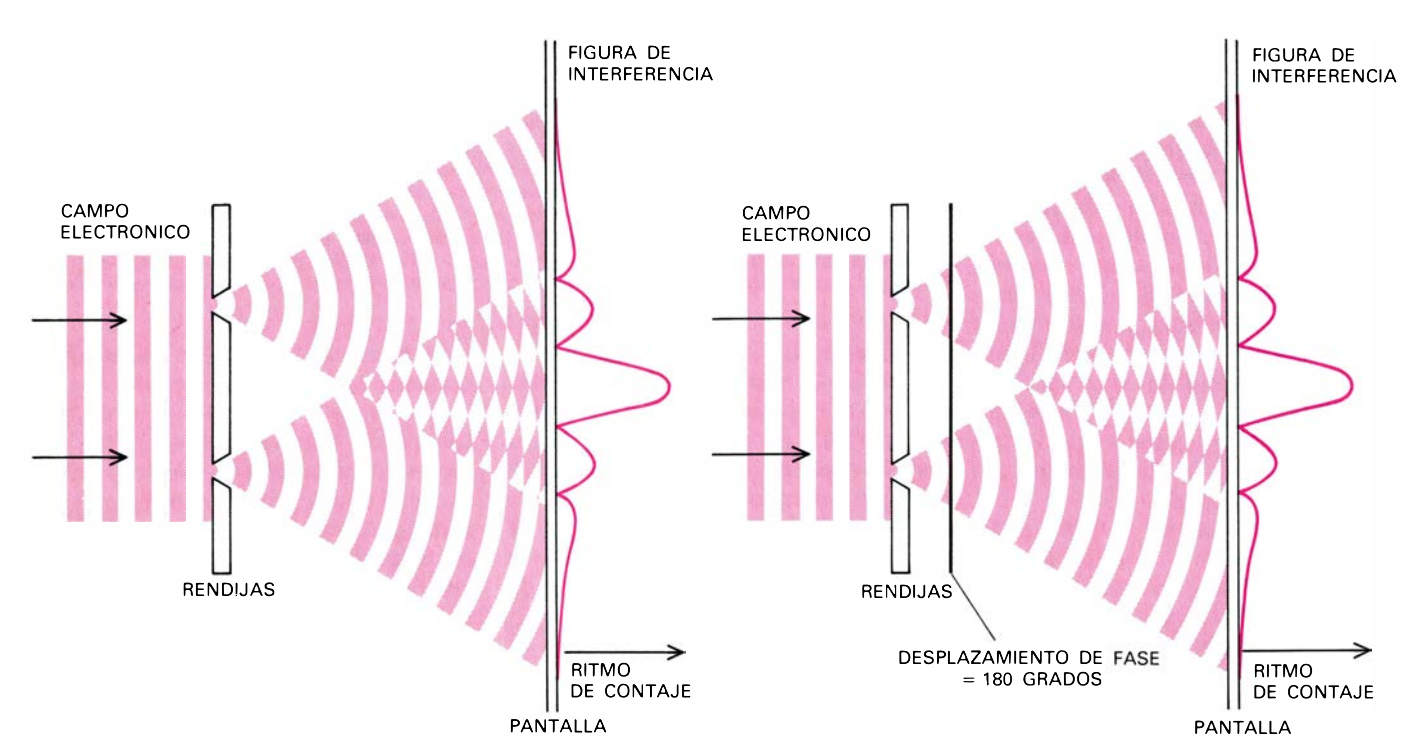
En mecánica cuántica, las propias partículas pueden representarse incluso como campos. Así, podemos considerar el electrón como un paquete de ondas dotado de una extensión finita en el espacio. A la inversa, muchas veces será oportuno representar un campo mecánico-cuántico como si se tratara de una partícula. La interacción de dos partículas a través de los campos que originan puede resumirse diciendo que ambas partículas intercambian una tercera partícula, el llamado cuanto del campo. Cuando dos electrones, rodeados cada uno de ellos por un campo electromagnético, se lanzan uno sobre otro y se repelen, se dice que intercambian un fotón, el cuanto del campo electromagnético.

El cuanto intercambiado tiene sólo una existencia efímera. Una vez emitido, debe ser reabsorbido, por la misma partícula o por otra, en un intervalo de tiempo finito. No puede existir indefinidamente, ni detectarse en ningún experimento. Entidades de esta clase son llamadas partículas virtuales. Cuanto mayor es su energía, más breve es su existencia. En efecto, una partícula virtual toma prestada o malversa una cantidad de energía, pero debe pagar la deuda antes de que pueda descubrirse el desfalco.

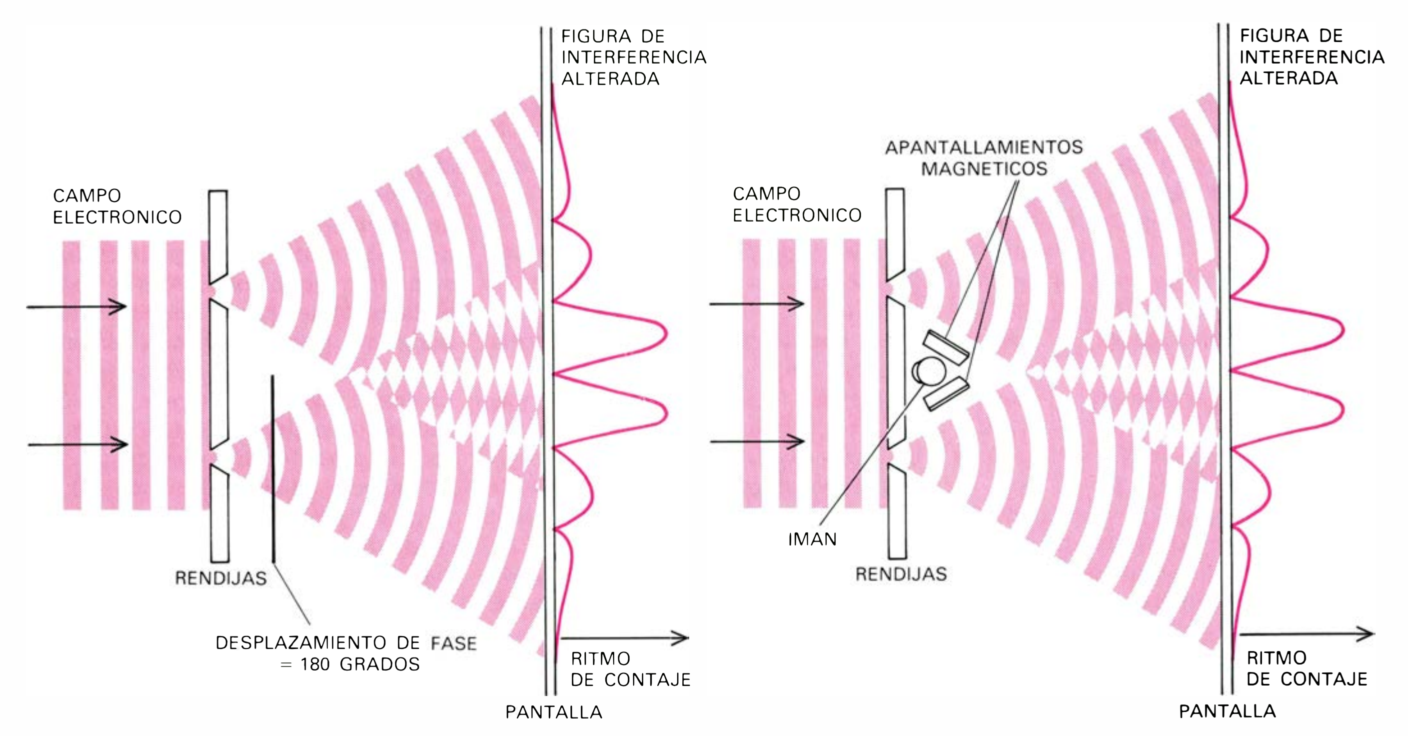
El alcance de la interacción está relacionado con la masa del cuanto intercambiado. Si el cuanto del campo posee una gran masa, debe pedir prestada más energía para poder sobrevivir, deuda que habrá de saldar antes para que no se descubra la discrepancia. La distancia que puede viajar la partícula antes de su reabsorción queda, por tanto, reducida y, en consecuencia, la fuerza correspondiente será de menor alcance. En el caso particular en que el cuanto intercambiado no tenga masa el alcance es infinito.

El número de componentes de un campo corresponde al número de estados mecánico-cuánticos de los cuantos del campo. A su vez, el número de estados posibles está relacionado con el momento angular intrínseco de spin de la partícula. El momento angular de spin puede tomar sólo valores discretos; cuando la magnitud del spin se mide en unidades fundamentales, es siempre un entero o semientero. Más aún, no es sólo la magnitud del spin lo que está cuantificado, sino también su dirección u orientación. (Precisando, el spin puede definirse como un vector paralelo a su eje de spin, y las proyecciones, o componentes, de este vector a lo largo de cualquier dirección del espacio deben tener valores que sean enteros o semienteros). El número de orientaciones posibles, o estados de spin, es igual a dos veces la magnitud del spin más uno. Así, una partícula de spin un medio, por ejemplo el electrón, posee dos estados de spin: el spin puede ponerse paralelo o antiparalelo a la dirección del movimiento de la partícula. Una partícula de spin uno tiene tres orientaciones, a saber, paralela, antiparalela y transversal. Una partícula de spin cero carece de eje de spin; como todas las orientaciones son equivalentes, se dice que tiene sólo un estado de spin.

Un campo escalar, dotado tan sólo de una componente (una magnitud), debe venir representado por un campo cuántico que tenga sólo una componente, o, en otras palabras, por una partícula de spin cero. Tales partículas se llaman, por consiguiente, partículas escalares. Similarmente, un campo vectorial de tres componentes exige que los cuantos del campo tengan un spin uno, con tres estados de spin: una partícula vectorial. El campo electromagnético es un campo vectorial, y el fotón, de acuerdo con estas especificaciones, posee spin unidad. Un campo



SIMETRIA GAUGE DEL ELECTROMAGNETISMO. Se trata de una invariancia con relación a los desplazamientos de fase del campo de materia que representa el electrón. La fase en sí no puede medirse, pero deja sentir su destacada influencia sobre cantidades observables tales como la figura de interferencia que se forma cuando las ondas de un campo electrónico atraviesan un par de rendijas. Los picos de esta figura se encuentran siempre que las ondas están en fase, y los nodos surgen cuando están en oposición de fase. Un desplazamiento en la fase altera totalmente la configuración del campo, pero sin cambiar la figura de interferencia observable. La simetría es exacta, por lo que el desplazamiento de fase no puede detectarse. De ahí que sea pura cuestión convencional la elección de la fase en cualquier descripción teórica del campo. En ausencia de fuerzas (o interacciones) que actúen entre electrones, sin embargo, la simetría será global: la figura observada será invariante sólo si el mismo desplazamiento de fase se realiza por doquier.



SIMETRIA DE GAUGE LOCAL del campo de materia del electrón restablecida al tener en cuenta los campos magnéticos. Si se desplaza la fase de uno de los haces de electrones difractados, pero no la del otro, se altera la figura de interferencia (*diagrama de la izquierda*). Sin embargo, puede obtenerse el mismo efecto al introducir un pequeño campo magnético perpendicular al haz del electrón y entre las rendijas (*diagrama de la derecha*). Nótese que el campo magnético induce un desplazamiento de fase aun cuando se coloquen apantallamientos de suerte que el campo no logre penetrar en la región donde las ondas del electrón se propagan e interfieren. Un experimentador que examine las figuras de interferencia no podrá distinguir entre los efectos de un desplazamiento de fase impuesto arbitrariamente sobre el haz del electrón y los efectos de un campo magnético introducido entre las rendijas. Cualquier desplazamiento local en la fase del campo material del electrón puede reproducirse por campos eléctricos y magnéticos; la fase del campo del electrón será arbitraria. La electrodinámica relaciona los campos de materia del electrón con los eléctricos y magnéticos.

gravitacional es una estructura más complicada llamada un tensor y tiene 10 componentes; sin embargo, no todas ellas son independientes y el cuanto del campo, el gravitón, tiene un spin de dos unidades, que de ordinario corresponde a cinco estados de spin.

En los casos del electromagnetismo y de la gravitación hay que considerar una nueva complicación. Puesto que el fotón y el gravitón carecen de masa, se deben mover siempre con la velocidad de la luz. Debido a su velocidad presentan una propiedad, no compartida por las partículas de masa finita: los estados de spin transversales no existen. Aunque en algún estado formal el fotón tiene tres estados de spin y el gravitón cinco, en la práctica sólo se pueden detectar dos de los estados de spin.

La primera teoría gauge con simetría local fue la teoría de los campos eléctricos y magnéticos introducida en 1868 por James Clerk Maxwell. La teoría de Maxwell se funda en la afirmación de que una carga eléctrica está rodeada por un campo eléctrico que se extiende hasta el infinito, y que el movimiento de una carga eléctrica da origen a un campo magnético que tiene también extensión infinita. Ambos campos son cantidades vectoriales y están definidos en cada punto del espacio por una magnitud y una dirección.

En la teoría de Maxwell, el valor del campo eléctrico en un punto viene determinado por la distribución de cargas alrededor de ese punto. Frecuentemente, conviene definir un potencial, o voltaje, determinado asimismo por la distribución de cargas: cuanto mayor sea la densidad de cargas en una región, más alto será su potencial. El campo eléctrico entre dos puntos viene dado por la diferencia de potencial entre ellos.

Para ilustrar el carácter de la simetría que hace que la teoría de Maxwell sea una teoría gauge podemos recurrir a un experimento imaginario. Supongamos un sistema de cargas eléctricas en un laboratorio, e imaginemos que se mida el campo electromagnético generado por estas cargas y que se tabulen las propiedades de este último. Si las cargas son estacionarias, no puede haber campo magnético (ya que el campo magnético se origina por el movimiento de una carga eléctrica); el campo será puramente eléctrico. En esta situación experimental se nota fácilmente la simetría global. La transformación de simetría consiste en elevar todo el laboratorio a un voltaje más alto, o, en otras palabras, a un potencial eléctrico más elevado. Si se repiten las mediciones no se observará ningún cambio en el campo eléctrico. La razón de ello estriba en que el campo, de acuerdo con la definición de Maxwell, viene determinado exclusivamente por las diferencias del potencial eléctrico y no por el valor absoluto del potencial. Por ese motivo, una ardilla puede pasearse sin peligro por un tendido de alta tensión no aislado.

Esta propiedad de la teoría de Maxwell corresponde a una simetría: el campo eléctrico es invariante con respecto a la suma o a la resta de un potencial constante arbitrario. Como se ha dicho antes, la simetría es de tipo global ya que el resultado del experimento permanece inalterable sólo si el potencial se cambia por doquier al mismo tiempo. Si se aumentara el potencial en una región y no en otra, cualquier experimento que cruzara la frontera vendría afectado por la diferencia de potencial, al igual que la ardilla sufre la descarga si toca simultáneamente una línea de alta tensión y un conductor conectado a tierra.

Pero una teoría cabal de los campos electromagnéticos ha de afrontar no sólo las distribuciones estáticas de car-

gas sino también las cargas en movimiento. A ese fin, la simetría global de la teoría debe convertirse en una simetría local. Si el campo eléctrico fuera el único que actuara entre partículas cargadas, no habría una simetría local. En realidad, cuando las cargas están en movimiento (y sólo entonces), el campo eléctrico no es el único presente: el propio movimiento origina un segundo campo, el campo magnético. Son los efectos del campo magnético los que restauran la simetría local.

Del mismo modo que el campo eléctrico, que depende de la distribución de las cargas, puede derivarse de forma útil de un potencial eléctrico, el campo magnético, que viene generado por el movimiento de las cargas, se describe más fácilmente como el resultado de un potencial magnético. Es en este sistema de campos de potencial donde las transformaciones locales pueden llevarse a cabo sin alterar en absoluto los campos eléctricos y magnéticos originales. El sistema de campos duales e interconectados tiene una simetría local exacta, de la que no goza el campo eléctrico por sí solo. Cualquier cambio local en el potencial eléctrico puede combinarse con un cambio compensador en el potencial magnético, de forma que los campos eléctricos y magnéticos sean invariantes.

La teoría de Maxwell del electromagnetismo es clásica, no mecánico-cuántica. Pero se puede probar que existe una simetría similar en la teoría cuántica de las interacciones electromagnéticas. En esa teoría, hay que describir el electrón como una onda o un campo, convención que en mecánica cuántica puede adoptarse para toda partícula material. Y resulta que en la teoría cuántica de los electrones un cambio en el potencial eléctrico implica un cambio en la fase de la onda del electrón.

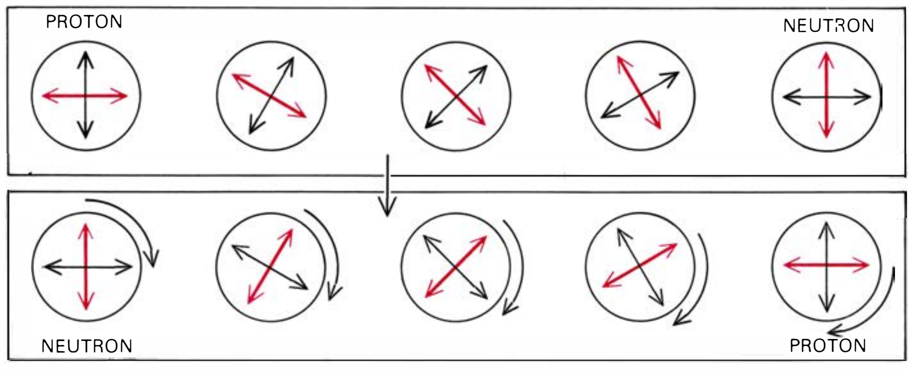
El electrón tiene spin de media unidad; posee, por tanto, dos estados de spin (paralelo y antiparalelo). En consecuencia, el campo asociado debe tener dos componentes. Cada una de las componentes debe venir representada por un número complejo, esto es, un número que tiene tanto una parte real, u ordinaria, y una parte imaginaria, que incluye como factor la raíz cuadrada de -1. El campo del electrón es un paquete de ondas en movimiento, que son oscilaciones en las amplitudes de las componentes reales e imaginarias del campo. Es necesario hacer notar que este campo no es el campo eléctrico creado por el electrón, sino un campo de materia. Existiría aunque el electrón careciera de carga eléctrica. Lo que el campo define es la probabilidad de hallar un electrón en un estado de spin especificado en un punto y en un instante dados. La probabilidad viene determinada por la suma de los cuadrados de las partes reales e imaginarias del campo.

En ausencia de campos electromagnéticos, la frecuencia de las oscilaciones en el campo del electrón es proporcional a la energía del electrón y la longitud de onda de las oscilaciones es inversamente proporcional a su momento. Para definir de forma completa las oscilaciones debe conocerse una cantidad adicional: la fase. La fase mide el desplazamiento de la onda desde un punto de referencia arbitrario y se expresa usualmente como un ángulo. Si en algún punto la parte real de la oscilación, por ejemplo, tiene una amplitud máxima positiva, a la fase en este punto se le puede asignar el valor de cero grados. Cuando a continuación la parte real se anula, la fase es de 90 grados y donde alcanza su máximo valor negativo la fase es de 180 grados. En general, la parte imaginaria de la amplitud está desfasada en 90 grados en relación con la parte real, de forma que cuando una parte tiene un máximo la otra se anula.

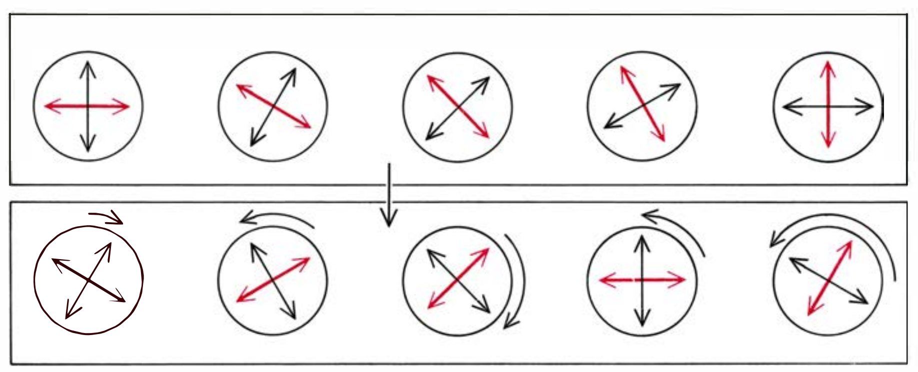
La única manera de determinar la fase de un campo electrónico consiste pues en separar las contribuciones de las partes real e imaginaria de la amplitud. Pero ello resulta imposible, incluso en línea de principio. La suma de los cuadrados de las partes reales e imaginarias puede ser conocida, pero no hay forma de determinar en un punto dado o en un instante determinado cómo contribuyen al total la parte real y la parte imaginaria. En realidad, una simetría exacta de la teoría implica que las dos contribuciones son indistinguibles. Puede medirse la diferencia en la fase del campo entre dos puntos o entre dos instantes, pero no la fase absoluta.

El hecho de que la fase de una onda electrónica no pueda medirse hace precisa una apostilla: la fase no puede

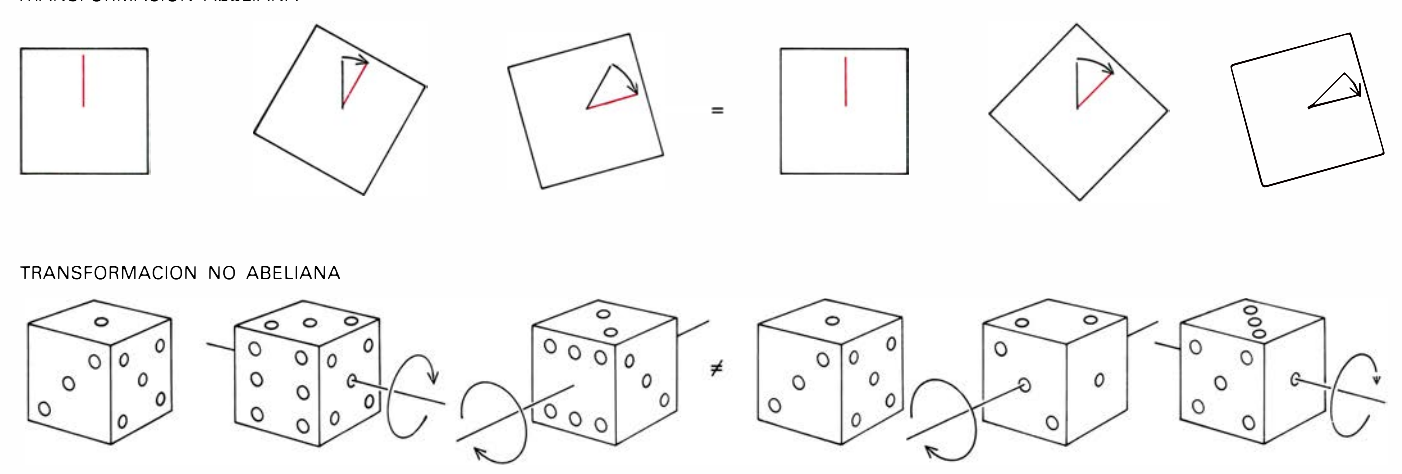
ROTACION GLOBAL DE SPIN ISOTOPICO



ROTACION LOCAL DE SPIN ISOTOPICO



SIMETRIA DE SPIN ISOTOPICO. Sirve de base a otra teoría gauge, discutida por vez primera en 1954 por C. N. Yang y Robert L. Mills. Si la simetría de spin isotópico es válida, la elección de qué posición de la flecha interna indica un protón y cuál indica un neutrón será pura cuestión convencional. La simetría global (*diagrama superior*) exige que se adopte la misma convención por doquier, y cualquier rotación de la flecha deberá hacerse de la misma forma en cualquier punto. En la teoría de Yang-Mills, la simetría de spin isotópico se convierte en una simetría local (*diagrama inferior*), y así se permite que la orientación de la flecha varíe de un lugar a otro. Para conservar la invariancia de todas las cantidades observables con respecto a tales transformaciones locales de spin isotópico, se exige introducir al menos seis campos, correspondientes a tres partículas vectoriales sin masa, o bosones vectoriales. Una de estas partículas puede identificarse con el fotón; las otras dos tienen carga eléctrica. La teoría ha gozado de cierto predicamento aunque, en su forma original, resultaba irrealista. Efectivamente, impedía la distinción entre protones y neutrones y predecía la existencia de partículas cargadas sin masa que no existen.



LOS EFECTOS DE LAS TRANSFORMACIONES repetidas distinguen la electrodinámica cuántica, que es una teoría abeliana, de la teoría de Yang-Mills, que es no abeliana. La transformación abeliana es conmutativa: si se aplican dos transformaciones sucesivas, el resultado es el mismo con independencia del orden en que se hagan. Tenemos un ejemplo de ello en la rotación en dos dimensiones. Las transformaciones no abelianas no son conmutativas, lo que significa que dos transformaciones dan resultados distintos, en general, si se cambia su orden. Las rotaciones en tres dimensiones presentan esta dependencia del orden. La electrodinámica cuántica es abeliana, pues los desplazamientos de fase sucesivos pueden aplicarse a un campo del electrón sin tener en cuenta el orden. La teoría de Yang-Mills es no abeliana, toda vez que el efecto neto de dos rotaciones de spin isotópico es generalmente distinto cuando se invierte la secuencia de las rotaciones. En efecto, una secuencia puede dar origen a un protón y la opuesta a un neutrón.

influir en el resultado de ningún experimento. Si lo hiciera, este experimento podría usarse para determinar la fase. Por tanto, el campo del electrón exhibe una simetría con relación a los cambios arbitrarios de fase. Añádase o sustráigase del campo electrónico cualquier ángulo: los resultados de todos los experimentos permanecerán inalterados.

Un ejemplo nos hará más inteligible ese principio: el experimento de difracción de electrones a través de dos rendijas, que constituye la demostración más clara de la naturaleza ondulatoria de la materia. Se hace pasar un haz de electrones a través de dos rendijas estrechas de una pantalla y se mide el número de electrones que alcanzan una segunda pantalla. La distribución de los electrones en la segunda pantalla origina una figura de difracción en la que se alternan los picos y los valles.

Para la mecánica cuántica, este experimento revela que la onda electrónica se divide en dos partes al llegar a la primera pantalla y las dos ondas difractadas interfieren a continuación la una con la otra. Cuando las ondas están en fase, la interferencia es constructiva y se cuentan muchos electrones en la segunda pantalla; cuando las dos ondas están en oposición de fase la interferencia destructiva reduce el contaje. Escuetamente: sólo la diferencia de fase determina la figura de interferencia. Si las fases de ambas ondas se desplazaran en la misma cantidad, la diferencia de fase de cada punto no sufriría alteración y observaríamos la misma figura de interferencia.

Son simetrías de esta clase, donde puede ajustarse arbitrariamente la fase del campo cuántico, las que reciben el nombre de simetrías gauge. Aunque el valor absoluto de la fase es irrelevante para los resultados de los experimentos, al construir una teoría de los electrones es aún necesario especificar la fase. La elección de un valor particular se llama una convención del gauge.

Simetría de gauge no es un término muy descriptivo para tal invariancia, pero el nombre tiene una larga historia y ya no se puede cambiar. Fue introducido alrededor de 1920 por Hermann Weyl, que andaba por aquel entonces preocupado por la formulación de una teoría que combinara el electromagnetismo con la teoría de la relatividad general. Weyl propuso una teoría que permaneciera invariante con respecto a dilataciones o contracciones arbitrarias del espacio. En esta teoría debía adoptarse un patrón distinto de longitud y de tiempo en cada punto del espacio-tiempo. El comparaba la elección de una convención de escalas con una elección de una galga, bloque de acero pulimentado usado por los maquinistas como patrones de longitud. La teoría, casi correcta, sólo necesitaba reemplazar "escalas de longitud" por "ángulos de fase". En su trabajo en alemán, Weyl habló de "Eich Invarianz", que se tradujo inicialmente por "calibration invariance" (invariancia de calibrado), pero la traducción alternativa de "gauge" se ha impuesto. (En castellano, la mayoría de los autores usan la palabra "gauge", por más que algunos prefieren traducir el término y hablan entonces de "contraste" o "aforo". N. T.)

La simetría del campo de materia del electrón descrita antes es una simetría global: la fase del campo debe variar de la misma forma por doquier en un instante. Se puede demostrar fácilmente que una teoría con sólo campos electrónicos, sin otras formas de materia o de radiación, no es invariante con relación a la correspondiente transformación de gauge local. Consideremos de nuevo el experimento de difracción de electrones por dos rendijas. Se procede como antes y se registra la figura de difracción de electrones. Se repite de nuevo el experimento, ajustando ahora a una de las rendijas el equivalente electro-óptico de una lámina de media onda, un aparato que cambia la fase de una onda en 180 grados. Cuando las ondas que se producen en las dos rendijas interfieren, la diferencia de fase entre ellas quedará alterada en 180 grados. Por cuya razón, en las situaciones en que antes la interferencia era constructiva será ahora destructiva, y viceversa. La figura de difracción observada no se mantendrá invariante; antes bien, se intercambiarán las posiciones de todos los picos y valles.

Supongamos que alguien deseara que la teoría fuera coherente con una simetría de gauge local. Quizá hubiera forma de lograrlo; en particular, a lo mejor se podría añadir otro campo que compensara los cambios en la fase del electrón. El nuevo campo habría de ser capaz, por supuesto, de hacer algo más que arreglar los resultados del experimento. Debería conservar la invariancia de todas la cantidades observables

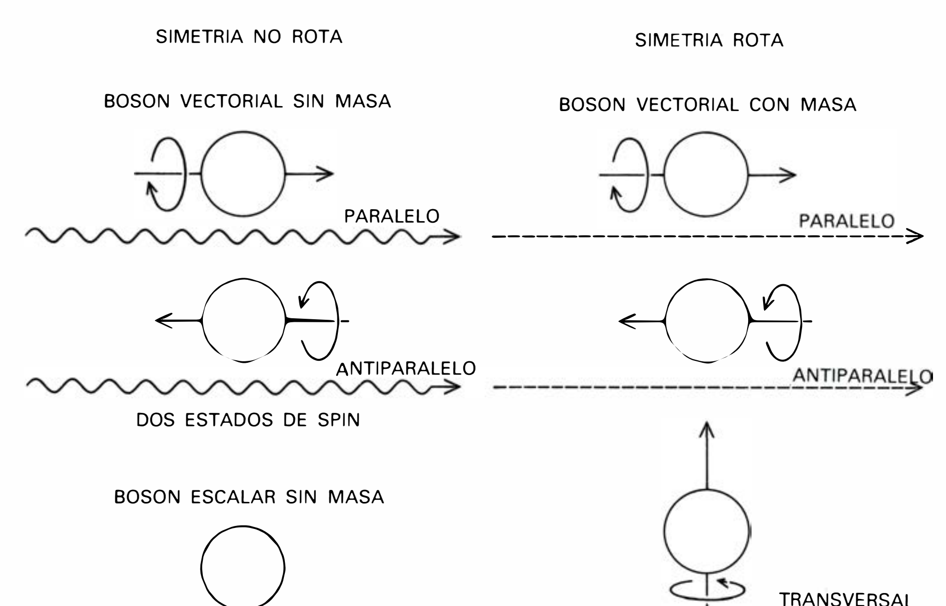
cuando la fase del campo electrónico sufriera cualquier alteración de un punto a otro y de un instante a otro. Matemáticamente, debiera permitirse que el desplazamiento de fase variase como una función arbitraria de posición y tiempo.

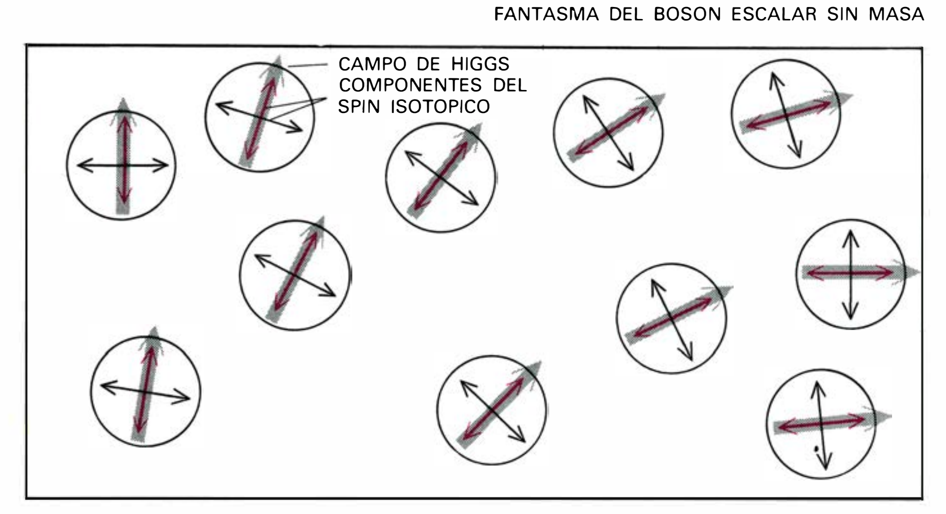
Por improbable que parezca, puede construirse un campo que cumpla estas especificaciones. El campo necesario resulta ser vectorial, correspondiente a cuantos del campo de spin una unidad. Más aún, el campo debe tener alcance infinito, ya que no hay límite a la distancia sobre la que las fases de los campos electrónicos deban ajustarse. La exigencia de un alcance infinito implica que los cuantos del campo hayan de poseer masa nula. Y esas propiedades se predican ya de un campo que nos es familiar: el campo electromagnético, cuyo cuanto es el fotón.

¿De qué forma el campo electromagnético asegura la invariancia gauge del campo electrónico? Recordemos que el campo electromagnético tiene por efecto transmitir fuerzas entre las partículas cargadas. Estas fuerzas pueden alterar el estado de movimiento de las partículas; lo que aún es más importante en este contexto, pueden alterar también la fase. Cuando un electrón absorbe o emite un fotón, la fase del campo del electrón cambia. Se demostró antes que el campo electromagnético por sí solo presenta una simetría local exacta; al describir los dos campos conjuntamente, la simetría local puede extenderse a ambos.

La conexión entre los dos campos yace en la interacción entre la carga del electrón y el campo electromagnético. Debido a esta interacción, la propagación de una onda de materia electrónica en un campo eléctrico puede describirse de forma adecuada sólo si se especifica el potencial eléctrico. Análogamente, para describir un electrón en un campo magnético se ha de especificar el potencial vector magnético. Cuando a estos dos potenciales se les asignan valores definidos, la fase de la onda del electrón queda fijada por doquier. Sin embargo, la simetría local del electromagnetismo permite dar al potencial eléctrico cualquier valor arbitrario, que puede elegirse independientemente en cada punto y en cada instante. Por esta razón, la fase del campo de materia del electrón puede tomar también cualquier valor en cualquier punto, sin que esta fase deje nunca de ser coherente con la convención adoptada para los potenciales eléctrico y magnético.

¿Qué tiene que ver esto con el experimento de difracción con dos rendijas? Pues que los efectos de un desplazamiento arbitrario en la fase de la onda del electrón pueden siempre imitarse aplicando un campo electromagnético. Así, el cambio observado en la figura de interferencia por haber interpuesto una placa de media onda delante de una de las rendijas, podría haberse logrado también de haber colocado las dos rendijas entre los polos de un imán. A partir de la figura resultante nadie podría adivinar cuál de los dos métodos se ha usado. Como las convenciones de gauge para los potenciales eléctrico y magnético pueden elegirse localmente, puede escogerse del mismo modo la fase del campo del electrón.





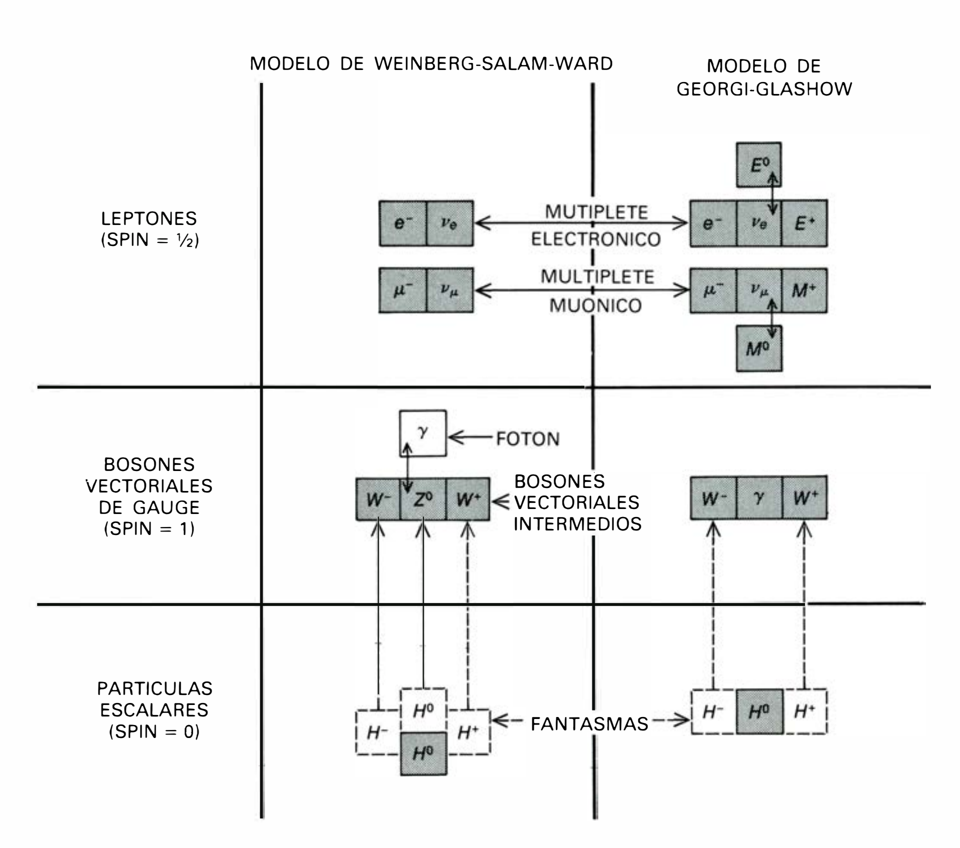
MECANISMO DE HIGGS, que permite dar masa a bosones vectoriales análogos al fotón de la teoría de Yang-Mills, con lo que la teoría se hace más realista. Los bosones sin masa tienen tres orientaciones posibles del spin (paralelo, antiparalelo y transversal a la dirección del movimiento); sólo dos de ellas son observables. El estado transversal no existe, peculiaridad ésta de las partículas sin masa, que se mueven a la velocidad de la luz. Si las partículas de Yang-Mills adquirieran masa, se observaría el estado transversal, y este nuevo modo de movimiento debería tener algún origen. En el mecanismo de Higgs, la fuente constituye un campo escalar adicional, correspondiente a un bosón de spin cero sin masa. Se dice que la partícula de Yang-Mills se "come" el bosón de Higgs, que a continuación se queda como un "fantasma" no observable. El campo de Higgs también proporciona un marco de referencia (*flechas grises*) en el que pueden distinguirse los protones de los neutrones. La flecha del campo de Higgs gira junto con las otras flechas en una transformación de gauge; no existe, por tanto, una orientación absoluta, si bien la orientación relativa de las flechas de spin isotópico pueden medirse con respecto a la flecha de Higgs. La simetría de la teoría, que en ausencia del mecanismo de Higgs había abolido todas las diferencias entre el protón y el neutrón, no se ha perdido; sólo se ha escondido.

La teoría que resulta de combinar los campos de materia del electrón con los campos electromagnéticos se llama electrodinámica cuántica. La formulación de la teoría y la demostración de su coherencia interna fue una labor que llevó unos 20 años; iniciada en la década de 1920 por P.A.M. Dirac, fue completada en lo esencial, alrededor de 1948, por Richard P. Feynman, Julian Schwinger, Sin-itiro Tomonaga y otros.

Las propiedades de simetría de la electrodinámica cuántica son sin duda atrayentes, pero la teoría sólo tendrá sentido físico si está de acuerdo con los resultados de la experimentación. En realidad, antes de poder realizar predicciones experimentales con sentido, la teoría debe pasar algunas pruebas de coherencia interna. Por ejemplo, las teorías mecánico-cuánticas predicen las probabilidades de los sucesos; las probabilidades no deben ser negativas y la suma de todas las probabilidades ha de ser 1. Además, las energías han de poseer valores positivos, pero no deben ser infinitas.

No se veía muy claro en un primer momento que la electrodinámica cuántica pudiera establecerse como una teoría física aceptable. En todos los intentos por calcular el resultado de hasta las más simples de las interacciones electromagnéticas, la interacción entre dos electrones por ejemplo, aparecía siempre un obstáculo. La secuencia más probable de sucesos en tal encuentro es que un fotón sea emitido por un electrón y que el otro electrón lo absorba. Pero caben también otros muchos intercambios más complicados; en realidad, el número de posibilidades es infinito. La probabilidad total de la interacción viene determinada por la suma de las contribuciones de todos los sucesos posibles.

Feynman introdujo un procedimiento sistemático para tabular estas contribuciones diseñando diagramas de los sucesos en una dimensión espacial y otra temporal. Particularmente molesta es aquella clase de diagramas que comprende "lazos" tales como el lazo en el espacio-tiempo, que se forma cuando se emite un fotón virtual y después lo reabsorbe el mismo electrón. Como mostramos antes, la energía máxima de una partícula está limitada sólo por el tiempo necesario para alcanzar su destino. Cuando un fotón es emitido y reabsorbido por la misma partícula, la distancia recorrida y el tiempo exigido pueden reducirse a cero, y, por tanto, la energía máxima



MODELO DE WEINBERG-SALAM-WARD, que incorpora el electromagnetismo y las fuerzas débiles en una teoría gauge local. El modelo se aplica a las interacciones de las partículas llamadas leptones, que incluyen el electrón (e^-), el muón (μ^-) y dos clases de neutrinos (ν_e y ν_μ). Al imponer que las interacciones de estas partículas permanezcan invariantes con respecto a las transformaciones locales de un equivalente leptónico del spin isotópico se originan cuatro campos sin masa. A continuación se da masa a tres de estos campos usando el mecanismo de Higgs; éstos son los bosones vectoriales intermedios W^+, W^- y Z^0. El cuarto bosón vectorial es el fotón. Tres de las partículas de Higgs son comidas por los bosones vectoriales y se transforman en fantasmas, pero la cuarta permanece en la teoría y debería ser observable. La teoría no unifica realmente las fuerzas electromagnéticas y las fuerzas débiles porque el fotón es aún por sí solo una familia. Una teoría propuesta por Howard Georgi y Sheldon Lee Glashow sugería una unificación más profunda, en la que el fotón y los bosones vectoriales con masa pertenecerían a la misma familia, pero dicha teoría está hoy día en contradicción con los datos experimentales.

puede ser infinita. Por esta razón algunos diagramas con lazos pueden dar una contribución infinita a la intensidad de la interacción.

Los infinitos encontrados en la electrodinámica cuántica condujeron inicialmente a predicciones que no tenían interpretación razonable entendidos como cantidades físicas. Cada interacción de electrones y fotones tenía asignada una probabilidad infinita. Los infinitos destruían incluso la descripción de un electrón aislado: como el electrón puede emitir y reabsorber partículas virtuales tiene una masa infinita y una carga infinita.

El remedio contra esta plaga de infinitos es el procedimiento llamado de renormalización. A grandes rasgos opera buscando un infinito negativo por cada infinito positivo, de forma que en la suma de todas las contribuciones posibles los infinitos se cancelen. Mérito de Schwinger y de los otros físicos que trabajaron en el problema fue llegar a demostrar que podía obtenerse un residuo finito por este método. El residuo finito es la predicción de la teoría. Queda determinado de forma única al exigir que todas las probabilidades de interacción resulten finitas y positivas.

La razón fundamental de este procedimiento puede explicarse de la forma siguiente. Cuando hacemos una medida en un electrón, lo que se mide realmente no es la masa o la carga de la partícula puntual con la que se inicia la teoría, sino las propiedades de un electrón juntamente con la nube de partículas virtuales que le rodea. Sólo se exige que la masa y la carga neta, las cantidades medibles, sean finitas en todos los estadios del cálculo. Las propiedades de un objeto puntual, que son llamadas la masa y la carga "desnuda", no están bien definidas.

Parecía en un comienzo como si hubiera de asignarse un valor infinito negativo a la masa desnuda, un absurdo que hizo que muchos físicos recelaran de la teoría renormalizada. Sin embargo, análisis más cuidadosos han demostrado que si la masa desnuda

tiene algún valor definido, éste tiende hacia cero. En cualquier caso, todas las cantidades con valores no plausibles son inobservables, incluso en principio. Hay otra objeción más profunda a la teoría: la electrodinámica cuántica no es perfecta desde un punto de vista matemático. Debido a los métodos que se emplean para hacer predicciones en la teoría, las predicciones quedan limitadas a una exactitud finita de algunos centenares de cifras decimales.

Evidentemente, la lógica y la coherencia interna del método de renormalización no acaban de satisfacer. Quizá la mejor defensa de la teoría sea simplemente decir que produce excelentes resultados. Ha conducido a resultados que están de acuerdo con los experimentos con la exactitud de una parte en mil millones, lo cual convierte a la electrodinámica cuántica en la teoría física más precisa de todas las que se han ideado. Es el modelo para las teorías de las otras fuerzas fundamentales y el patrón con el que se comparan tales teorías.

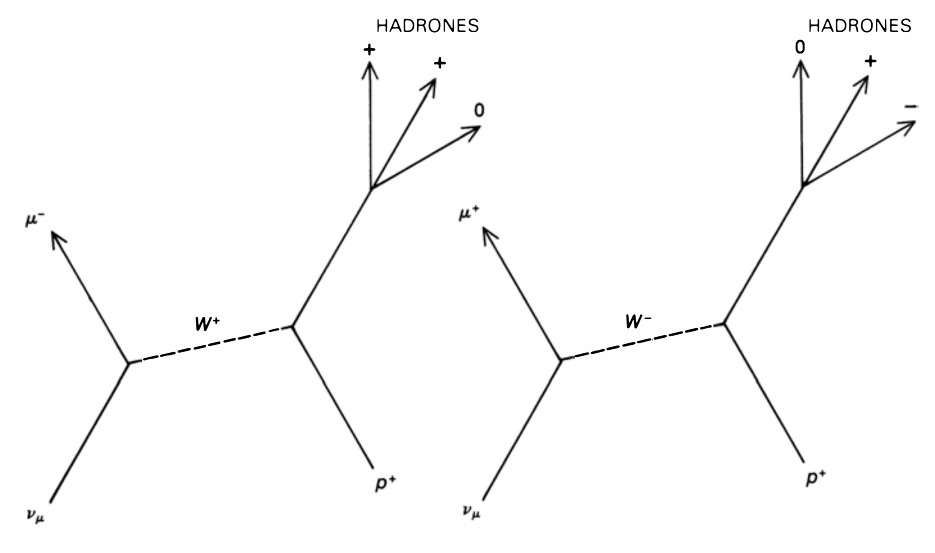
Cuando se llegó a la elaboración final de la teoría de la electrodinámica cuántica, existía, desde hacía ya unos 30 años, otra teoría basada en una simetría gauge local. Se trataba de la teoría de la relatividad general de Einstein. Esta simetría no era la de un campo distribuido a través del espacio y del tiempo, sino de la misma estructura del espacio-tiempo.

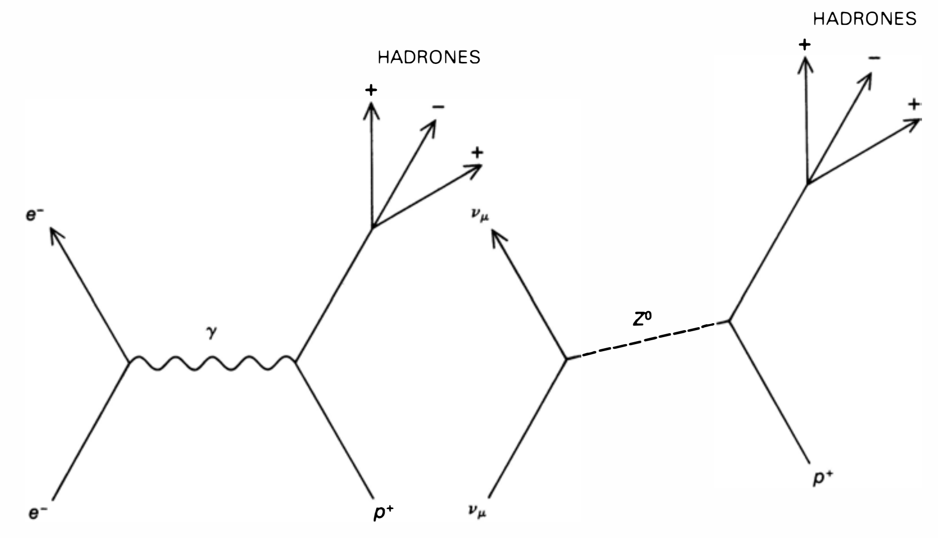
Cada punto del espacio-tiempo puede caracterizarse mediante cuatro números, que dan su posición en las tres dimensiones espaciales y su secuencia en la dimensión temporal. Estos números son las coordenadas de un suceso, y el procedimiento usado para asignar tales números a cada punto del espacio-tiempo se conoce por sistema de coordenadas. En la Tierra, por ejemplo, las tres coordenadas espaciales usuales son la longitud, la latitud y la altura; la coordenada temporal puede ser dada en horas contadas a partir del mediodía. En este sistema de coordenadas, el origen, punto donde las cuatro coordenadas tienen valor nulo, está al mediodía al nivel del mar en el punto donde el meridiano de Greenwich cruza el ecuador.

La elección de tal sistema de coordenadas es evidentemente fruto de un acuerdo. Los barcos podrían navegar lo mismo si el origen del sistema de coordenadas estuviera desplazado a Utrecht, Holanda. A cada punto de la Tierra y a cada suceso en su historia tendrían que asignarse nuevas coordenadas, pero los cálculos realizados con ellas conducirían invariablemente a los mismos resultados que los cálculos hechos en el sistema inicial. En particular, cualquier cálculo de la distancia entre dos puntos daría el mismo resultado.

La libertad de mover el origen de un sistema de coordenadas constituye una simetría de la naturaleza. Hay tres simetrías que están relacionadas: todas las leyes de la naturaleza permanecen invariantes cuando transformamos el sistema de coordenadas por una traslación, por una rotación o por una reflexión especular. Importa sobremanera no perder de vista que tales simetrías son sólo globales. Cada transformación de simetría se puede definir como una fórmula para encontrar las nuevas coordenadas de un punto a partir de las antiguas. Estas fórmulas deben aplicarse simultáneamente de la misma forma a todos los puntos.

La teoría general de la relatividad nace de la observación fundamental de que la estructura del espacio-tiempo no es necesariamente coherente con un sistema de coordenadas construido en su integridad por líneas rectas cortándose en ángulos rectos; antes bien cabe la posibilidad de exigirse un sistema de coordenadas curvilíneo. Las líneas de longitud y latitud sobre la Tierra consti-





CORRIENTES DEBILES NEUTRAS. Constituyen la prueba decisiva del modelo de Weinberg-Salam-Ward. Se creía antes que todas las interacciones débiles implicaban una transferencia de carga eléctrica entre las partículas que interaccionaban; estos sucesos podrían explicarse con sólo dos bosones vectoriales intermedios, el W^+ y el W^-. Sucesos en los que no había transferencia de carga eran característicos de las interacciones electromagnéticas, donde la partícula virtual intercambiada era un fotón. El modelo de Weinberg-Salam-Ward predice que las interacciones débiles pueden ocurrir también sin intercambio de carga; estas corrientes débiles neutras vienen mediadas por el bosón neutro Z^0, que es idéntico al fotón pero tiene una masa grande. Las corrientes débiles neutras se observaron por primera vez en 1973.

tuyen un sistema de esta clase, puesto que ellas siguen la curvatura de la Tierra.

En tal sistema es fácil imaginar una transformación local de coordenadas. Supongamos que la altura fuera definida como la distancia vertical a la superficie terrestre, en lugar de definirse sobre el nivel medio del mar. Al abrir un pozo, alteraríamos el sistema de coordenadas, pero sólo para aquellos puntos que se hallaran directamente sobre el pozo. La acción excavadora representa la transformación local de coordenadas. Parece como si las leyes de la física (o las normas de navegación) no permanecieran invariantes después de tales transformaciones; en un universo libre de fuerzas gravitacionales ello sería así. Un avión inducido a volar a una altura constante descendería de repente cuando pasara sobre la excavación, y las aceleraciones necesarias para seguir el nuevo perfil del terreno serían fácilmente detectables.

Como en electrodinámica, la simetría local puede restablecerse con sólo añadir un nuevo campo a la teoría; en la relatividad general el nuevo campo es por supuesto el de la gravitación. La presencia de este nuevo campo ofrece una explicación alternativa de las aceleraciones detectadas en el avión, podrían ser consecuencia, no de un cambio local en el sistema de coordenadas, sino de una anomalía en el campo gravitacional. No importa la fuente de la anomalía: podría obedecer a una concentración de masa en la Tierra o un objeto distante en el espacio. Lo interesante es que puede reproducirse cualquier transformación local del sistema de coordenadas mediante un conjunto apropiado de campos gravitacionales. El piloto del avión no podría distinguir un efecto del otro.

La teoría del electromagnetismo de Maxwell y la teoría de la gravitación de Einstein deben mucho de su belleza a una simetría de gauge local; sus éxitos han sido desde hace mucho tiempo una fuente de inspiración para los físicos teóricos. En cambio, las explicaciones teóricas de las otras dos fuerzas de la naturaleza resultaron menos satisfactorias, hasta hace poco. Una teoría de las fuerzas débiles formulada en los años treinta por Enrico Fermi explicaba algunos de los hechos básicos de las interacciones débiles, pero la teoría carecía de simetría local. Las interacciones fuertes parecían una jungla de campos misteriosos y de partículas resonantes. Se ha desvelado ahora la razón por la que costó tanto entender el sentido de todas esas fuerzas: no se comprendía la teoría gauge local necesaria para ello.

C. N. Yang y Robert L. Mills, que estaban entonces en el Brookhaven National Laboratory, dieron el primer paso teórico en 1954. Independientemente y al mismo tiempo, R. Shaw, de la Universidad de Cambridge, avanzó una propuesta muy parecida. Inspiradas por el éxito de las otras teorías gauge, estas teorías partían de una simetría global ya existente y se preguntaban qué consecuencias tendría si se la hiciera local.

La simetría considerada en la teoría de Yang-Mills es la simetría de spin isotópico, la regla que establece que las interacciones fuertes de la materia permanecen invariantes (o casi invariantes) cuando se intercambian protones por neutrones. En la simetría global, cualquier rotación de las flechas internas que indican el spin isotópico debe ser llevada a cabo simultáneamente por doquier. Postulando una simetría local se puede variar independientemente la orientación de las flechas de un lugar a otro y de un instante a otro. Las rotaciones de las flechas pueden depender de cualquier función arbitraria de la posición y el tiempo. Esta libertad de elegir distintas convenciones para la identidad de una partícula nuclear es una simetría de gauge local.

Como en otros casos donde una simetría global se convierte en una local, la invariancia sólo puede mantenerse si se añade algo más a la teorías. Por ser la teoría de Yang-Mills más complicada que las teorías gauge anteriores, resulta que debe añadirse mucho más. Cuando las rotaciones de spin isotópico se hacen arbitrarias de un lugar a otro, las leyes de la física permanecen invariantes sólo si se añaden seis nuevos campos. Todos ellos son campos vectoriales y todos tienen un alcance infinito.

Los campos de Yang-Mills se construyen tomando como modelo el electromagnetismo; a decir verdad, dos de ellos pueden identificarse con los campos usuales eléctricos y magnéticos. En otras palabras, describen el campo de un fotón. Los restantes campos de Yang-Mills también pueden ser agrupados en pares e interpretarse como campos eléctricos y magnéticos, pero los fotones que describen difieren en un aspecto crucial de las propiedades conocidas del fotón: son aún partículas sin masa de spin uno, pero tienen carga eléctrica. Uno es negativo y el otro es positivo.

El imponer una carga eléctrica a un fotón tiene notables consecuencias. El fotón se define como el cuanto del campo que conduce las fuerzas electromagnéticas de una partícula a otra. Si el propio fotón tiene carga, pueden existir interacciones electromagnéticas directas entre los fotones. Por citar un ejemplo, dos fotones con cargas opuestas podrían quedar ligados formando un átomo de "luz". El fotón neutro usual nunca interacciona consigo mismo de esta forma.

Los sorprendentes efectos de los fotones cargados se aprecian mejor cuando se aplica una transformación de simetría local más de una vez a la misma partícula. En la electrodinámica cuántica, como indicábamos antes, la operación de simetría es un cambio local en la fase del campo del electrón, de suerte que cada uno de tales cambios de fase viene acompañado por una interacción con el campo electromagnético. Es fácil imaginar un electrón que sufre dos desplazamientos de fase sucesivos: emitiendo un fotón y absorbiendo después otro. La intuición sugiere que si se invirtiera la secuencia de los desplazamientos de fase, de modo que primero fuera absorbido un fotón y más tarde fuera emitido otro, el resultado final sería el mismo. Y esto es lo que realmente sucede. Se puede realizar un número ilimitado de desplazamientos de fase: el resultado final será simplemente la suma algebraica de todos los desplazamientos, con independencia de su secuencia.

En la teoría de Yang-Mills, donde la operación de simetría es una rotación local de la flecha de spin isotópico, el resultado de transformaciones múltiples puede diferir totalmente. Supongamos que un hadrón sufre una transformación de gauge, *A*, seguida poco después por una segunda transformación, *B;* al final de esta secuencia, se observará que la flecha de spin isotópico está en la orientación que corresponde a un protón. Supongamos ahora que se aplicaran las mismas transformaciones al mismo hadrón en el orden inverso: *B* seguida de *A*. En general el estado final no será el mismo; la partícula puede ser un neutrón en lugar de un protón. El efecto neto de las transformaciones depende explícitamente de la secuencia en que hayan sido realizadas.

En razón de esa diferencia se dice que la electrodinámica cuántica es una teoría abeliana y la teoría de Yang-Mills, una teoría no abeliana. La terminología se toma de la teoría de grupos y

honra la memoria de Niels Henrik Abel, matemático noruego que vivió en los primeros años del siglo XIX. Los grupos abelianos están compuestos de transformaciones que, aplicándolas una después de la otra, tienen la propiedad conmutativa; los grupos no abelianos no son conmutativos.

La conmutación es una propiedad familiar de la aritmética; se predica de la suma y la multiplicación, donde para cualesquiera números A y B podemos decir que $A + B = B + A$ y $A \times B = B \times A$. Para indicar de qué manera el concepto puede aplicarse a un grupo de transformaciones recurramos a un ejemplo conocido: el grupo de las rotaciones. Todas las rotaciones posibles de un objeto bidimensional son conmutativas; por tanto, el grupo de estas rotaciones es abeliano. Por ejemplo, rotaciones de + 60 grados y −90 grados dan una rotación neta de −30 grados con independencia de cuál se haya aplicado en primer lugar. Para un objeto tridimensional libre de girar alrededor de tres ejes la ley conmutativa no es válida, y el grupo de rotaciones tridimensionales no es abeliano. Como ejemplo, consideremos un avión que se dirija al norte en vuelo nivelado. Si se desvía 90 grados hacia la izquierda y a continuación gira a la izquierda 90 grados alrededor de su eje longitudinal, el avión acaba dirigiéndose hacia el oeste y con el extremo del ala izquierda dirigida hacia abajo. Cambiando la secuencia de las transformaciones de forma que el giro de 90 grados venga seguido por el desvío de 90 grados, coloca el avión en un descenso en picado con las alas alineadas en la dirección norte-sur.

Igual que la teoría de Yang-Mills, la teoría de la relatividad general es no abeliana: al hacer dos transformaciones sucesivas de coordenadas, el orden en el que se realizan influye, en general, en el resultado. En los últimos 10 años, más o menos, se han ideado otras varias teorías no abelianas, y hasta las propias interacciones electromagnéticas han sido incorporadas en una teoría mayor que es no abeliana. Al menos por ahora parece como si todas las fuerzas de la naturaleza vinieran gobernadas por teorías gauge no abelianas.

La teoría de Yang-Mills ha resultado tener una importancia monumental, aunque su formulación inicial era totalmente inapropiada para describir el mundo real. Una primera objeción es que en ella la simetría de spin isotópico es exacta, lo que comporta que los neutrones y los protones sean indistinguibles; situación que, obviamente, contradicen los hechos. Más preocupante todavía es la predicción de fotones cargados eléctricamente. La masa del fotón debe ser necesariamente nula pues ha de tener un alcance infinito. La existencia de cualquier partícula cargada eléctricamente más ligera que el electrón alteraría el mundo hasta el punto de no reconocerlo. Por supuesto, no se ha observado una tal partícula. A pesar de estas dificultades, la teoría es muy bella y atrayente desde un punto de vista filosófico. Para obviar esas dificultades se adoptó la estratagema de dotar artificialmente a los cuantos del campo cargado con una masa mayor que cero.

Los cuantos de los campos cargados no desaparecen por imponerles una masa, aunque sí se les confina en un alcance finito. Si la masa es suficientemente grande, el alcance puede achicarse cuanto se desee. A medida que los efectos de largo alcance se vayan eliminando, la existencia de tales campos podría hacerse compatible con los hechos experimentales. Más aún, la selección del campo de Yang-Mills neutro como el único real de largo alcance distingue automáticamente los protones de los neutrones. Por ser este campo el electromagnético, distinguiremos el protón del neutrón en virtud de su diferente interacción con él; en otras palabras, por sus distintas cargas eléctricas.

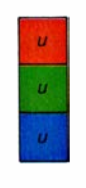
Con esta modificación, la simetría local de la teoría de Yang-Mills no sería exacta, sino sólo aproximada, ya que la rotación de la flecha del spin isotópico tendría ahora consecuencias observables. Pero no se trata de ninguna objeción fundamental: las simetrías aproximadas son muy comunes en la naturaleza. (La simetría bilateral del cuerpo humano sólo es aproximada). Más aún, a distancias mucho más pequeñas que el alcance de las componentes con masa del campo de Yang-Mills, la simetría local va ganando en precisión. Así pues, en cierto sentido, la estructura microscópica de la teoría podría permanecer localmente simétrica, pero no sus predicciones de sucesos macroscópicos, observables.

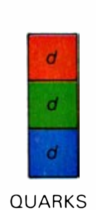
La teoría modificada de Yang-Mills se hacía más inteligible, pero era aún necesario darle una interpretación mecánico-cuántica. El problema de los infinitos resultó ser más difícil de lo que había sido en la electrodinámica cuántica y la receta normal de la renormalización no lo resolvía. Había que hallar nuevas técnicas.

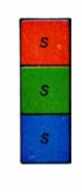
Feynman aportó, en 1963, una idea valiosa: la noción de partícula "fantasma", una partícula añadida a la teoría en el transcurso del cálculo que desaparece cuando el cálculo ha finalizado. Se sabe desde el principio que la partícula fantasma es ficticia, pero se puede justificar su uso si nunca aparece en el estado final, lo que puede asegurarse haciendo que la probabilidad total de producir una partícula fantasma sea siempre cero.

De entre los grupos teóricos que continuaban trabajando en la teoría de

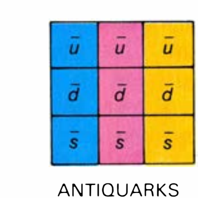
FERMIONES (SPIN = ½)







QUARKS



ANTIQUARKS

BOSONES VECTORIALES DE GAUGE (SPIN = 1)



GLUONES

PARTICULAS ESCALARES (SPIN = 0) NINGUNA

MEDIANTE EL MODELO QUARK se nos permite describir todos los hadrones, que engloban el protón y el neutrón, como partículas compuestas.Se hallarían constituidas por entidades más pequeñas, llamadas quarks. En su forma original, el modelo suponía que los quarks podrían tener tres "aromas", denominados *u*, *d* y *s* (del inglés "*up*", arriba, "*down*", abajo y "*strange*", extraño), cada uno de los cuales sabemos que puede tener tres "colores", a saber: rojo, verde y azul. Hay antiquarks con sus respectivos anticolores cian, magenta y amarillo. Las interacciones de los quarks se describen mediante una teoría gauge basada en la invariancia con respecto a las transformaciones locales de color. Se necesitan dieciséis campos para mantener esta invariancia. Se toman en pares para formar ocho bosones vectoriales sin masa, llamados gluones, dotado cada uno de ellos de una combinación de color y anticolor.

HADRON SIN COLOR

TRANSFORMACION GLOBAL DE COLOR

HADRON SIN COLOR

HADRON SIN COLOR

TRANSFORMACION LOCAL DE COLOR SIN FUERZAS

HADRON COLOREADO

HADRON SIN COLOR

TRANSFORMACION LOCAL DE COLOR CON FUERZAS FUERTES

HADRON SIN COLOR

GLUON

LA SIMETRIA DE COLOR exige que todos los hadrones permanezcan blancos, o sin color, aun cuando se altere el color de los quarks constituyentes. El color de un quark, igual que los estados de spin isotópico de los hadrones, puede venir indicado por la posición de una flecha en un espacio interno imaginario. La simetría global se obtiene fácilmente. Si un hadrón consta en un comienzo de tres quarks, uno de cada uno de los colores, cualquier rotación sincronizada de las tres flechas deberá dejar el balance general de los colores invariante. Ahora bien, en la ausencia de fuerzas entre quarks, la simetría global no puede convertirse en una simetría local. Si se cambia la posición de una flecha de color dejando las otras dos fijas, el hadrón adquiere un color neto. Para conservar la simetría local del color, hay que introducir fuerzas. En particular, cuando se cambia el color de un quark, se emitirá una partícula virtual que habrá de reajustar los colores de los otros quarks al objeto de que el hadrón como un todo no tenga color. Los campos necesarios para asegurar que todos los hadrones sean no coloreados forman los ocho campos gluónicos de la cromodinámica cuántica.

Yang-Mills, el método de la partícula fantasma sólo se tomó en serio en la Universidad de Utrecht, donde yo cursaba entonces mi carrera. Martin J. G. Veltman, mi director de tesis, junto con John S. Bell, de la Organización Europea para la Investigación Nuclear (CERN), Ginebra, llegaron a la conclusión de que las interacciones débiles podían describirse por alguna forma variante de la teoría de Yang-Mills. Veltman emprendió un análisis sistemático del problema de la renormalización en el modelo de Yang-Mills modificado (con campos cargados con masa), examinando una a una todas las clases de diagramas de Feynman. Demostró fácilmente que los diagramas que no tenían lazos cerrados sólo originaban contribuciones finitas a la probabilidad total de interacción. Los diagramas con un lazo originaban términos infinitos, pero al sacarle partido a las propiedades de las partículas fantasma se logró que los infinitos positivos y los negativos se cancelaran exactamente.

A medida que el número de lazos aumentaba, se elevaba rápidamente el número de diagramas. Más aún, se volvían más complejos los cálculos necesarios para cada diagrama. Al objeto de subvenir a la enorme tarea de comprobar todos los diagramas con dos lazos se escribió un programa de ordenador que operaba con todas las manipulaciones algebraicas de las probabilidades. La salida del programa fue un elenco de los coeficientes de las cantidades infinitas que quedaban después de que las contribuciones de todos los diagramas se hubieran sumado. Si se eliminaban los infinitos de la teoría, los coeficientes deberían ser nulos sin excepción. En 1970 se conocieron ya los resultados y se había excluido la posibilidad de un error. Quedaban unos cuantos infinitos.

El fallo de la teoría de Yang-Mills modificada no debía imputarse a ningún defecto de la propia formulación de Yang-Mills, sino a sus modificaciones. Las masas de los campos cargados habían de introducirse "a mano", de donde resultaba que la invariancia con respecto a las rotaciones de spin isotópico locales no era totalmente perfecta. Por aquel entonces, los investigadores rusos L. D. Faddeev, V. N. Popov, E. S. Fradkin y I. V. Tyutin sugirieron que una teoría de Yang-Mills pura, con sólo campos sin masa, podría ser renormalizable. La dificultad con que tropieza esta teoría estriba en que, amén de ser irreal, tiene campos de largo alcance con los que no es fácil trabajar.

Entre tanto, Robert H. Brout, de la Universidad de Bruselas, y Peter Higgs, de la de Edimburgo, habían introducido un nuevo elemento en la formulación de las teorías gauge. Dieron con un método para dotar de masa a algunos campos de Yang-Mills y que al mismo tiempo no rompiera la simetría exacta. Dicha técnica se conoce hoy por mecanismo de Higgs.

La idea fundamental del mecanismo

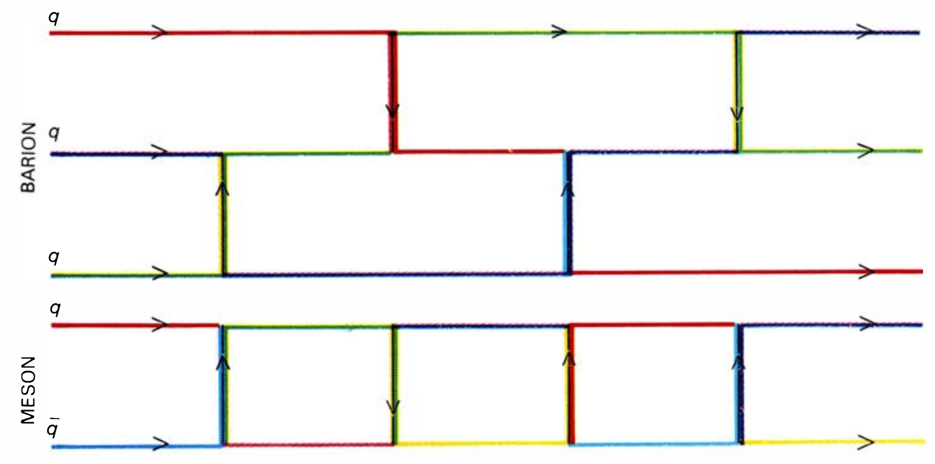
de Higgs consiste en incluir en la teoría un campo extra, un campo que tenga la propiedad peculiar de no anularse en el vacío. Solemos imaginar el vacío como aquel espacio donde no hay nada; en física, sin embargo, el vacío se define de forma más precisa: un estado en el que todos los campos tienen los menores valores posibles de energía. Para muchos campos, la energía es mínima cuando el valor del campo es cero por doquier o, en otras palabras, cuando el campo está "desconectado". Un campo electrónico, por ejemplo, tiene su energía mínima cuando no hay electrones. El campo de Higgs resulta paradójico en este punto. Reducirlo a cero cuesta energía; la energía del campo es menor cuando el campo tiene algún valor uniforme mayor que cero.

El efecto del campo de Higgs es dar un sistema de referencia en el que es posible determinar la flecha del spin isotópico. El campo de Higgs puede representarse como una flecha superpuesta a los otros indicadores de spin isotópico en un espacio interno imaginario del hadrón. Lo que distingue la flecha del campo de Higgs es que tiene una longitud fija, que viene determinada por el valor de campo en el vacío. La orientación de las otras flechas de spin isotópico puede entonces medirse con respecto al eje definido por el campo de Higgs. De esta forma un protón puede distinguirse de un neutrón.

Pudiera parecer que la introducción del campo de Higgs debería alterar la simetría gauge de la teoría y conducir de nuevo a infinitos no eliminables. La realidad es que la simetría gauge no queda destruida sino simplemente escondida. La simetría especifica que todas las leyes de la física deben permanecer invariantes cuando la flecha de spin isotópico se gira de una forma arbitraria de un sitio a otro. Esto implica que la orientación absoluta de la flecha no puede ser determinada, puesto que cualquier experimento para medir la orientación debería detectar alguna variación en una cantidad física cuando la flecha se girara. Con la inclusión del campo de Higgs, no puede determinarse todavía la orientación absoluta de la flecha porque la flecha que representa el campo de Higgs también gira durante una transformación de gauge. Todo lo que se puede medir es el ángulo entre la flecha del campo de Higgs y las otras flechas de spin isotópico, o, en otras palabras, sus orientaciones relativas.

El mecanismo de Higgs constituye un ejemplo de un proceso conocido por rotura espontánea de una simetría, que ya estaba bien establecido en otras áreas de la física. El concepto fue introducido por Werner Heisenberg en su descripción de los materiales ferromagnéticos. Heisenberg destacó que la teoría que describe un ferromagneto tiene una simetría geométrica perfecta y que no establece ninguna dirección privilegiada en el espacio. Sin embargo, cuando el material queda magnetizado, hay un eje —la dirección de magnetización— que puede distinguirse de entre todos los demás. La teoría es simétrica pero el objeto que describe no lo es. Análogamente, la teoría de Yang-Mills retiene su simetría gauge con respecto a las rotaciones de la flecha de spin isotópico, pero los objetos descritos —protones y neutrones— no presentan dicha simetría.

¿De qué manera el mecanismo de Higgs puede dar masa a los cuantos del campo de Yang-Mills? El procedimiento puede explicarse de la siguiente forma. El campo de Higgs es una cantidad escalar, que tiene una sola magnitud; por tanto, el cuanto del campo deberá tener spin cero. Los campos de Yang-Mills son vectores, como el campo electromagnético, y vienen representados por cuantos de spin uno. Normalmente, una partícula con spin de una unidad tiene tres estados de spin (orientado paralelamente, antiparalelamente y transversalmente a su dirección del movimiento); ahora bien, puesto que las partículas de Yang-Mills carecen de masa y se mueven con la velocidad de la luz constituyen un caso especial; sus estados transversales no existen. Si las partículas adquirieran masa, perderían entonces esta situación privilegiada y deberían ser observables los tres estados de spin. En mecánica cuántica, el equilibrio de estados de spin es riguroso y el estado extra debe venir de algún sitio: procede del campo de Higgs; a raíz de ello cada cuanto de Yang-Mills gana masa y un estado de spin, en tanto que la partícula de Higgs desaparece. Sugerida por Abdus Salam, del Centro Internacional de Física Teórica en Trieste, tenemos una descripción pintoresca de este proceso: las partículas de Yang-Mills sin masa se "comen" las partículas de Higgs para ganar peso, y las partículas de Higgs que han sido tragadas se convierten en fantasmas.



INTERCAMBIO DE GLUONES que mantiene un barión (formado por tres quarks) o un mesón (formado por un quark y un antiquark) incoloros. En este proceso se conserva el color total de las partículas. Así, un quark rojo puede convertirse en un quark verde con sólo emitir un gluon que tenga el color rojo y el anticolor magenta: el magenta puede interpretarse como el antiverde. Por tanto, el rojo del quark se lo lleva el rojo del gluon, y el verde y el antiverde se crean en cantidades iguales. Si el gluon es absorbido por un quark, el verde del quark y el antiverde del gluon se aniquilarán entre sí. En virtud del continuo intercambio de los gluones no puede haber un cambio neto del color del hadrón.

Veltman me propuso, en 1971, investigar la renormalización de las teorías de Yang-Mills puras. Las reglas para construir los diagramas de Feynman necesarios habían sido ya formuladas por Faddeev, Popov, Fradkin y Tyutin e, independientemente, por Bryce S. De Witt, de la Universidad de Texas en Austin, y Stanley Mandelstam, de la Universidad de California en Berkeley. Parecía bueno adaptar a este trabajo los poderosos métodos para los estudios de renormalización que había desarrollado antes el mismo Veltman.

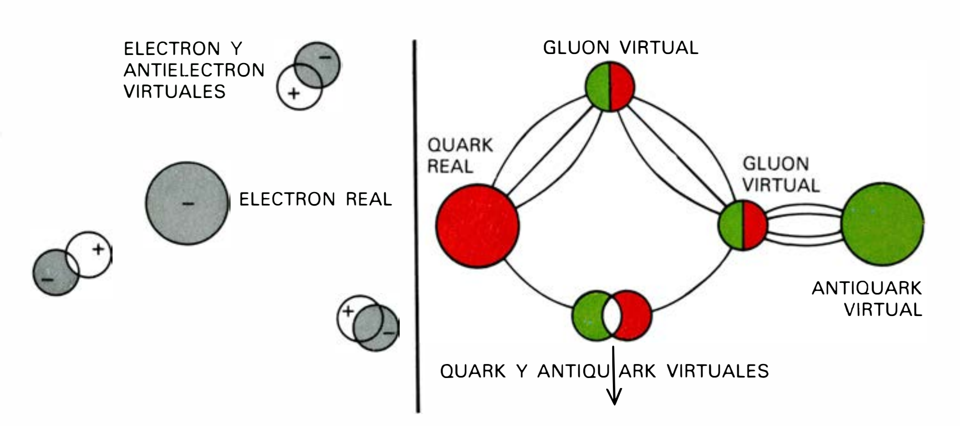
Desde el punto de vista formal, los resultados eran alentadores. Mas, si la teoría había de ser realista, era obligado hallar algún medio que confinara los campos de Yang-Mills a un alcance finito. Acababa de enterarme en un cursillo de verano de que Kurt Symanzih, del Sincrotrón de Electrones ale-

mán, y Benjamin W. Lee, del Fermi National Accelerator Laboratory estadounidense, habían logrado con éxito la renormalización de un modelo teórico en que se rompía espontáneamente una simetría global. Parecía natural intentar el mecanismo de Higgs en la teoría de Yang-Mills, donde la simetría rota es local.

Algunos modelos simples daban resultados alentadores: en los casos elegidos todos los infinitos se cancelaban con independencia del mínimo de partículas gauge que se intercambiasen y los lazos que se incluyeran en los diagramas de Feynman. La prueba decisiva vendría cuando a la teoría se le aplicara el programa de ordenador que aislaba infinitos en todos los diagramas posibles con dos lazos. Los resultados de esta prueba se obtuvieron en julio de 1971; la salida del programa era una sucesión ininterrumpida de ceros. Cada uno de los infinitos se cancelaba exactamente. Comprobaciones posteriores mostraron que los infinitos tampoco aparecían en diagramas de Feynman de suma complejidad. Mis resultados fueron pronto confirmados por otros, en particular por Lee y por Jean Zinn-Justin, del Centro de Investigaciones Nucleares de Saclay, en las cercanías de París.

La teoría de Yang-Mills empezó como un modelo de las interacciones fuertes, pero en la época en que fue renormalizada, el interés por ella se había centrado en sus aplicaciones a las interacciones débiles. En 1967, Steven Weinberg, de la Universidad de Harvard, e independientemente (aunque más tarde) Salam y John C. Ward, éste de la Universidad Johns Hopkins, habían propuesto un modelo de interacciones débiles basado en una versión de la teoría de Yang-Mills según la cual los cuantos de gauge adquirían masa por el mecanismo de Higgs. Supusieron que quizá fuera posible renormalizar la teoría, pero no lo demostraron. Su opinión se sumó así a muchas otras conjeturas no probadas hasta que, años más tarde, mis propios resultados demostraron que era precisamente aquella subclase de las teorías de Yang-Mills que incorporaban el mecanismo de Higgs las que podían ser renormalizadas.

La característica más notable de la fuerza débil es su corto alcance: su influencia eficaz se siente sólo a distancias de 10^{-15} centímetros, una centésima parte, aproximadamente, del radio del protón. La fuerza es débil principalmente por su alcance tan corto: es poco probable que las partículas se acerquen lo suficiente unas a otras para poder interaccionar. El corto alcance significa que las partículas virtuales intercambiadas en las interacciones débiles deben ser muy masivas. Las estimaciones actuales son de 80 a 100 veces la masa del protón.



POLARIZACION DEL VACIO que explica en parte la ley peculiar de fuerza que, según parece, permite que los quarks se muevan en completa libertad en el interior del hadrón y prohíbe que puedan aislarse quarks o gluones. En electrodinámica cuántica (*izquierda*), pares de electrones y antielectrones virtuales rodean cualquier carga aislada, tal como un electrón. Debido a las fuerzas electrostáticas, los antielectrones cargados positivamente tienden a permanecer más cerca de la carga negativa del electrón; cancelan, pues, parte de ella. La carga del electrón observada es la diferencia entre la carga "desnuda" y la carga apantallante de los antielectrones virtuales. De un modo semejante, pares virtuales de quarks disminuyen la intensidad de la fuerza entre un quark real y un antiquark real. En cromodinámica cuántica hallamos un efecto competitivo que no aparece en electrodinámica cuántica. Puesto que los gluones tienen también carga de color (en tanto que el fotón carece de carga eléctrica), los gluones virtuales influyen sobre la magnitud de la fuerza de color entre quarks. Los gluones no apantallan la carga del quark, sino que la refuerzan. Resulta así que la carga de color es débil y los quarks se mueven libremente mientras se encuentren cerca; distantes, puede exigirse una energía infinita para separarlos.

El modelo de Weinberg-Salam-Ward abarca en realidad tanto las fuerzas débiles como el electromagnetismo. La conjetura sobre la que el modelo descansa es un postulado de invariancia local con respecto al spin isotópico; para mantener la invariancia se introducen cuatro campos tipo fotón, en lugar de los tres de la teoría de Yang-Mills original. El cuarto fotón podría ser identificado con alguna forma primordial de electromagnetismo. Corresponde a una fuerza separada, que habría de añadirse a la teoría sin explicación. Por cuya razón, el modelo no debería ser llamado una teoría de campos unificada. Las fuerzas permanecen distintas; es la forma como se entremezclan lo que hace el modelo tan peculiar.

En un principio, los cuatro campos del modelo de Weinberg-Salam-Ward tienen alcance infinito y, por tanto, deben ser transportados por cuantos de masa nula; un campo lleva una carga eléctrica negativa, otro lleva una carga positiva y los otros dos campos son neutros. La rotura espontánea de la simetría introduce cuatro campos de Higgs, cada uno de ellos representado por una partícula escalar. Tres de los campos de Higgs son tragados por las partículas de Yang-Mills, de forma que las dos partículas de Yang-Mills cargadas y una de las neutras adquieren una gran masa. Estas partículas se denominan colectivamente bosones vectoriales intermedios con masa, y se denotan por W^+, W^- y Z^0. La cuarta partícula de Yang-Mills, que es neutra, permanece sin masa: es el fotón del electromagnetismo. De las partículas de Higgs, las tres que suministran masa a las partículas de Yang-Mills se transforman en fantasmas, y no pueden, por tanto, ser observadas, pero la última partícula no queda absorbida, y debería verse si se dispusiera de energía suficiente para su producción.

La predicción más intrigante del modelo era la existencia del Z^0, una partícula idéntica al fotón en todos los aspectos salvo en su masa, la cual no se había incluido en ninguna de las descripciones primitivas y provisionales de las fuerzas débiles. Sin el Z^0 cualquier interacción débil debería forzosamente implicar un intercambio de carga eléctrica. Sucesos de esta clase se llaman sucesos de corrientes débiles cargadas. El Z^0 introducía una nueva clase de interacciones débiles, sucesos debidos a corrientes débiles neutras. Al intercambiar un Z^0, las partículas interaccionarían sin ninguna transferencia de carga y podrían mantener sus identidades iniciales. Las corrientes débiles neutras fueron observadas por primera vez en el CERN en 1973.

La elaboración de una teoría correcta de las interacciones fuertes, que sólo sufren los hadrones, no podía acome-

terse mientras no se hubiera comprendido un punto fundamental relativo a los hadrones: éstos no constituyen partículas elementales. En 1963, Murray Yel-Mann, del Instituto de Tecnología de California, avanzó un modelo de hadrones entendidos como objetos compuestos. Con independencia, y contemporáneamente, propusieron ideas similares Yuval Ne'eman, de la Universidad de Tel Aviv, y George Zweig, del Cal Tech. En ese modelo, los hadrones están constituidos por partículas más pequeñas, que Gell-Mann llamó quarks. Pueden construirse hadrones a partir de quarks según dos métodos. Por combinación de tres quarks se obtiene un barión, una clase de hadrones que comprende el protón y el neutrón. Ligando un quark y un antiquark se obtiene un mesón, siendo los piones miembros típicos de esta clase. Cada uno de los hadrones conocidos puede explicarse como una combinación permitida de quarks.

En el modelo original había sólo tres clases de quarks, llamados "arriba" (up), "abajo" (down) y "extraño" (strange). James D. Bjorken, del Stanford Linear Accelerator Center, y Sheldon Lee Glashow, de Harvard, propusieron muy pronto añadir un cuarto quark portador de una propiedad llamada "encanto" (charm). En 1971 Glashow, John Iliopoulos, de París, y Luciano Maiani, de la Universidad de Roma, demostraron mediante un elegante razonamiento que era necesario un quark con encanto para salvar una discrepancia en la teoría gauge de las interacciones débiles. Concluyeron que los quarks con encanto debían existir si tanto la teoría gauge como la teoría de los quarks eran correctas. El descubrimiento en 1974 de la partícula *J* o psi, que consiste en el estado ligado de un quark y un antiquark encantados, corroboró el modelo de Weinberg-Salam-Ward y convenció a muchos físicos de que el modelo de quarks debía tomarse con seriedad en todos sus aspectos. Hoy día creemos que al menos son necesarios dos nuevos "aromas" o clases de quarks, que se les ha llamado "cima" (top) y "fondo" (bottom).

La primera tarea de cualquier teoría de las interacciones fuertes es explicar las reglas que permiten construir hadrones a partir de quarks. No entraña especial dificultad abordar la estructura del mesón; puesto que sólo consta de un quark y un antiquark, basta con suponer que los quarks tienen alguna propiedad análoga a la carga eléctrica. Que un quark y un antiquark se liguen vendría justificado por el principio de atracción de cargas opuestas, como ocurre en el átomo de hidrógeno. Enigma más arcano parece ser la estructura del barión. Para explicar de qué manera tres quarks pueden formar un estado ligado hay que presuponer que tres cargas del mismo tipo se atraen.

La teoría que se ha desarrollado para explicar las fuerzas fuertes prescribe exactamente esas interacciones. El análogo de la carga eléctrica es una propiedad llamada color (aunque no tiene nada que ver con los colores del espectro visible). Se eligió el término color porque las reglas de formación de hadrones pueden expresarse sucintamente exigiendo que todas las combinaciones permitidas de los quarks sean "blancas" o carezcan de color. A los quarks se les asignan los colores primarios rojo, verde y azul; los antiquarks tienen los "anticolores" complementarios cian, magenta y amarillo. Cada aroma de los quarks aparece en los tres colores, de suerte que la introducción de la carga de color triplica el número de quarks distintos.

A partir de los pigmentos de quarks disponibles se llega a dos maneras de crear el blanco: por mezcla de los tres colores primarios o haciendo lo propio con un color primario y un anticolor complementario. Los bariones se construyen de acuerdo con el primer esquema: los tres quarks del barión deben tener colores distintos, de modo que los tres pigmentos primarios estén necesariamente presentes. En un mesón, el color vendrá siempre acompañado por su anticolor complementario.

La teoría ideada para dar cuenta de estas sorprendentes interacciones recuerda, paso a paso, la electrodinámica cuántica, razón por la cual se la bautizó cromodinámica cuántica. Se trata de una teoría gauge no abeliana. La simetría gauge es una invariancia con respecto a las transformaciones locales del color de los quarks.

Es fácil imaginar una simetría global correspondiente al color. Los colores de los quarks, igual que los estados de spin isotópico de los hadrones, pueden indicarse por la orientación de una flecha en un espacio interno imaginario. Rotaciones sucesivas de un tercio de vuelta cambiarían un quark de rojo a verde y de nuevo a rojo. En un barión habrá tres flechas, una por cada color. Por definición, la transformación de simetría global debe afectar a las tres flechas de la misma forma y en el mismo instante. Supongamos que las tres flechas giren en el sentido de las agujas del reloj un tercio de vuelta. A raíz de tal transformación, los tres quarks cambiarán de color, pero todas las propiedades observables del hadrón permanecerán invariantes. En particular, continuará habiendo uno de cada color y, por tanto, el barión seguirá sin color.

La cromodinámica cuántica exige que se mantenga esta invariancia aun cuando la transformación de simetría sea local. En ausencia de fuerzas o interacciones se pierde la invariancia. Una transformación local puede cambiar el color de un quark y dejar inalterados los de los otros quarks, lo que daría un color neto al hadrón. Como en las otras teorías gauge, para restaurar la invariancia con respecto a las operaciones de simetría local hay que introducir nuevos campos. En cromodinámica, los campos necesarios son análogos al campo electromagnético, si bien mucho más complicados; tienen ocho veces tantas componentes cuantas ostenta el campo electromagnético. Esos campos originan la fuerza fuerte.

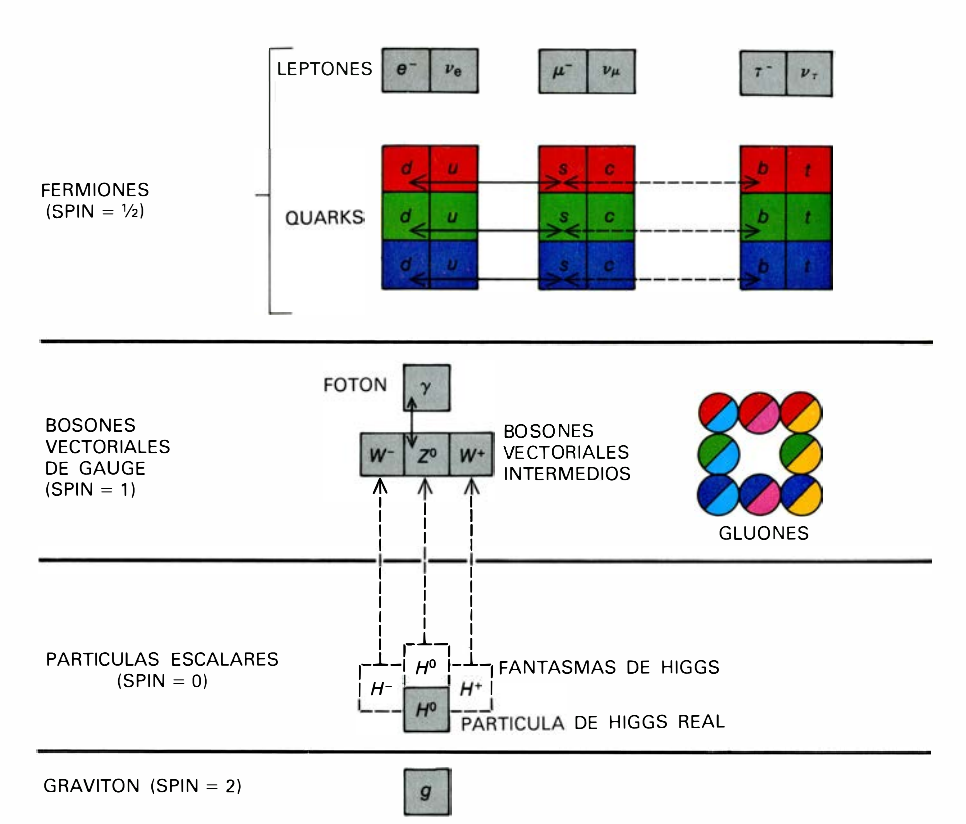
Los cuantos del campo de color se llaman gluones (porque pegan [en inglés, "glue"] los quarks entre sí). Hay ocho clases de ellos, todos con masa nula y con momento angular de spin de una unidad. En otras palabras se trata de bosones vectoriales carentes de masa, como el fotón. Asimismo, igual que el fotón, los gluones son eléctricamente neutros, pero no lo son con respecto al color. Cada gluon tiene un color y un anticolor. Hay nueve combinaciones posibles de color y anticolor; una de ella, sin embargo, es equivalente al blanco y se excluye, quedando ocho campos gluónicos distintos.

Los gluones conservan la simetría local de color. Y la mantienen de la siguiente forma. Un quark es libre de cambiar su color, y puede hacerlo independientemente de los otros quarks, pero cada transformación de color debe ir acompañada de la emisión de un gluon, al igual que el electrón puede cambiar su fase con sólo emitir un fotón. El gluon, que se propaga a la velocidad de la luz, es absorbido entonces por otro quark, que desplazará su color en exactamente la cantidad necesaria para compensar el cambio original. Supongamos, por ejemplo, que un quark rojo cambia su color a verde, y en el proceso se emite un gluon que tiene los colores rojo y antiverde. El gluon es absorbido por un quark verde, y, en la reacción que tiene lugar, el verde del quark y el antiverde del gluon se aniquilan entre sí, dejando el segundo quark de color rojo. Por tanto, en el

estado final, lo mismo que en el estado inicial, hay un quark rojo y un quark verde. A causa del continuo intercambio de los gluones no puede haber un cambio neto del color del hadrón, aunque los colores de los quarks varíen libremente de un punto a otro. Todos los hadrones permanecen blancos, y la fuerza fuerte no es nada más que el sistema de interacciones necesario para mantener esta condición.

A pesar de la complejidad de los campos gluónicos, la electrodinámica y la cromodinámica cuánticas se parecen mucho a nivel formal. Destaquemos que el fotón y el gluon son idénticos en su spin y en que no tienen ni carga ni masa. Resulta curioso, entonces, que las interacciones de los quarks sean tan distintas de las de los electrones.

Electrones y quarks forman estados ligados, es decir, átomos en el caso de los electrones y hadrones en el de los quarks. Los electrones se observan también como partículas independientes; una pequeña cantidad de energía es suficiente para aislar un electrón al ionizar un átomo. Nunca se ha encontrado un quark aislado. Parece que es imposible ionizar un hadrón por mucha energía que se le dé. Los quarks están ligados tan fuertemente que no se pueden separar. Pero lo más paradójico de todo reside en que los estudios de la estructura interna de los hadrones demuestran que los quarks se mueven con entera libertad como si no estuvieran ligados en absoluto.



MODELO STANDAR de las interacciones de las partículas elementales que describe las cuatro fuerzas de la naturaleza mediante tres teorías gauge no abelianas. Las partículas fundamentales de la materia son seis leptones y quarks de seis aromas distintos, pudiendo cada aroma presentar tres colores distintos. El electromagnetismo y la fuerza débil son mediados por las partículas de gauge del modelo de Weinberg-Salam-Ward, es decir, el fotón, que no tiene masa, y un triplete de bosones vectoriales de masa muy grande, los W^+, W^- y Z^0. La fuerza fuerte se atribuye a los ocho gluones sin masa de la cromodinámica cuántica. La gravitación resulta del intercambio de una partícula sin masa de spin dos, el gravitón, que viene descrito por otra teoría gauge local: la teoría general de la relatividad. Además queda una partícula de Higgs, que tiene masa y es eléctricamente neutra. En los próximos años la búsqueda de los bosones vectoriales con masa y de la partícula de Higgs podrán poner a prueba el valor de esta síntesis.

No se han observado gluones por vía experimental y de un modo directo. Su propia presencia en la teoría provoca objeciones parecidas a las formuladas contra la teoría de Yang-Mills pura, sin masas. Si existieran partículas sin masa que se parecieran tanto al fotón, deberían ser fáciles de detectar y se conocerían desde haría tiempo. Cabría dar a los gluones una masa a través del mecanismo de Higgs, pero se hace incómodo tener que ocultar ocho gluones por este procedimiento. Más aún, la masa debería ser elevada, pues de lo contrario los gluones ya habrían sido producidos en experimentos con aceleradores de alta energía; si la masa es grande resulta que el alcance de las fuerzas que ligan los quarks se hace demasiado pequeño.

Se ha dado con una solución provisional de este dilema al examinar, sin modificar, con mayor atención las propiedades de los campos de color. Al discutir la renormalización en electrodinámica cuántica hice notar que incluso un electrón aislado está rodeado de una nube de partículas virtuales, que él emite y reabsorbe continuamente. Entre las partículas virtuales no hay sólo partículas neutras, el fotón, por ejemplo, sino también pares de partículas de cargas opuestas, tales como electrones y antipartículas, los positrones. Son las partículas virtuales cargadas de esta nube las que en circunstancias ordinarias esconden la carga desnuda negativa e "infinita" del electrón. En la proximidad de la carga desnuda, los pares electrón positrón se polarizan ligeramente: los positrones virtuales, bajo la influencia atractiva de la carga desnuda, están, en promedio, más cercanos a ella que los electrones virtuales, que son repelidos. De ello resulta la neutralización parcial de la carga desnuda; lo que se ve a grandes distancias es la diferencia entre la carga desnuda y la carga apantallante de los positrones virtuales. Sólo cuando el cuerpo de prueba se acerca a menos de 10^{-10} centímetros empiezan a ser importantes los efectos del no apantallamiento de la carga desnuda.

Es razonable suponer que el mismo efecto debe ser operativo entre cargas de color, y así ocurre. Un quark rojo está envuelto por pares de quarks y antiquarks; las cargas antirrojas de esta nube son atraídas hacia el quark central y tienden a apantallar su carga. Ahora bien, existe en cromodinámica cuántica un efecto competitivo que no aparece en electrodinámica cuántica. Mientras que el fotón carece de carga eléctrica y por tanto no ejerce ningún influjo directo en el apantallamiento de los electrones, los gluones tienen carga de color. (Esta diferencia obedece a que la electrodinámica cuántica es una teoría abeliana, mientras que la cromodinámica cuántica es no abeliana.) Los pares de gluones virtuales forman una nube alrededor de un quark coloreado, pero resulta que los gluones tienden a incrementar el color en vez de atenuarlo. Es como si la componente roja del gluon fuera atraída hacia el quark rojo y añadiera por tanto su carga a la carga total efectiva. Si no hay más de 16 aromas de quarks (y por ahora sólo conocemos la existencia de seis), el "antiapantallamiento" de los gluones constituye el influjo dominante.

Se llega a ese curioso comportamiento de los gluones tras cálculos muy complicados; la interpretación de los

resultados depende además de cómo se hayan realizado los cómputos. Cuando los estimé, hallé que la fuerza responsable era el análogo en color del campo magnético del gluon. Importa también que los gluones virtuales puedan emitirse solos, mientras que los quarks virtuales aparecen siempre como un quark y un antiquark. Un único gluon, con una carga neta de color, incrementa la fuerza que actúa entre otras dos cargas de color.

En virtud de ese "antiapantallamiento", la fuerza de color efectiva de un quark aumentaba con la distancia. Un quark lejano reacciona ante los campos combinados del quark central y de las cargas gluónicas que lo refuerzan; a pequeñas distancias, una vez haya penetrado dentro de la nube gluónica, sólo se muestra eficaz la carga desnuda, más pequeña. Por tanto, los quarks de un hadrón actúan de forma parecida a como si estuvieran enganchados por tiras de goma: a distancias muy pequeñas, cuando las tiras están flojas, los quarks se mueven casi independientemente, pero a una distancia mayor, cuando las tiras se encuentran muy tensas, los quarks están fuertemente ligados.

La polarización de los gluones virtuales conduce a una explicación bastante precisa del comportamiento de los quarks a pequeñas distancias. Cuando la ligazón es débil, podemos calcular el movimiento esperado de las partículas con buena aproximación. La interacción de largo alcance, y en particular el hecho de que los quarks y gluones no aparezcan como partículas libres, cabe atribuirlo al mismo mecanismo de antiapantallamiento gluónico. Parece probable que cuando dos cargas de color tienden a separarse las fuerzas entre ellas crezcan indefinidamente, de suerte que precisaríamos una energía infinita para lograr una separación macroscópica. Este fenómeno de confinamiento permanente de los quarks puede estar relacionado con ciertas propiedades matemáticas peculiares de la teoría gauge. Obra en nuestro favor que algunos modelos teóricos muy simplificados incorporen un enfrentamiento permanente. Aunque en la teoría completa fallen todos los métodos de cálculo cuando las fuerzas se hacen muy grandes, el principio parece válido. Quarks y gluones se hallan, a buen seguro, permanentemente confinados en el interior de los hadrones.

Si la versión actual de la cromodinámica cuántica es correcta,la simetría de color constituye una simetría exacta y los colores de las partículas no pueden distinguirse en absoluto. Se trata de una teoría de gauge pura de la clase propuesta en un principio por Yang y Mills. Los campos de gauge, de largo alcance, recuerdan muchísimo, desde el punto de vista formal, el campo del fotón. Las ligaduras mecánico-cuánticas sobre estos campos son tan fuertes, sin embargo, que las interacciones observadas resultan totalmente distintas de las del electromagnetismo e incluso conducen al confinamiento de una clase entera de partículas.

No siempre allí donde los teorías gauge son correctas son útiles. Los cálculos exigidos para predecir el resultado de un experimento son tediosos, y, salvo en electromagnética cuántica, pocas veces podemos vanagloriarnos de haber obtenido una gran precisión. A razones prácticas o técnicas como éstas se debe que no se haya resuelto todavía el problema del confinamiento de los quarks. Las ecuaciones que describen un protón en términos de quarks y gluones son del mismo orden de complicación que las que describen un núcleo de tamaño intermedio en términos de protones y neutrones. Ninguno de estos conjuntos de ecuaciones puede resolverse con exactitud.

A pesar de estas limitaciones, las teorías gauge han contribuido enormemente al conocimiento de las partículas elementales y sus interacciones. No importa tanto el atractivo filosófico del principio de simetría local o el mismo éxito de teorías individuales, cuanto la creciente convicción de que esta clase de teorías que ahora se consideran incluyen todas las posibles teorías para cualquier sistema de partículas cuyas interacciones mutuas no sean demasiado fuertes. Los experimentos demuestran que si las partículas permanecen agrupadas a distancias menores de 10^{-14} centímetros, su interacción total, incluidos los efectos de todas las fuerzas conocidas o no, es realmente pequeña. (Los quarks son un caso particular: aunque las interacciones entre ellos no son pequeñas, las mismas pueden atribuirse a los efectos de las partículas virtuales, y sabido es que las interacciones de las partículas virtuales son sólo moderadas.) Por tanto, es razonable intentar un ajuste sistemático de las teorías gauge existentes con los datos experimentales.

Aunque el soporte matemático de las teorías gauge es rígido, queda libertad para el ajuste. La magnitud predicha de una interacción entre partículas no sólo depende de la estructura de la teoría sino también de los valores asignados a ciertos parámetros libres, que han de considerarse como constantes de la naturaleza. La teoría mantiene su coherencia con independencia de los valores que se asignen a estas constantes; las predicciones experimentales, en cambio, dependen fuertemente de los valores asignados a estas constantes. Aunque las constantes pueden medirse por vía experimental, nunca pueden derivarse de la teoría. Ejemplos de estas constantes de la naturaleza son la carga del electrón y las masas de las partículas elementales, tales como el electrón y los quarks.

El vigor de las teorías gauge radica en que se precisa un número relativamente pequeño de tales parámetros: bastan 18 constantes de la naturaleza para explicar todas las fuerzas conocidas. Los enmarañados fenómenos de las interacciones fuertes de las partículas, que parecían incomprensibles hace 15 años, pueden ser ahora desvelados mediante una teoría que tiene sólo un puñado de parámetros libres. Todos ellos, menos tres, son lo suficientemente pequeños para poderlos ignorar.

Aun cuando los parámetros libres han quedado reducidos a un número manejable, constituyen una parte esencial de la teoría. No podemos dar ninguna explicación de los valores que asumen. Las preguntas fundamentales que no han respondido aún las teorías gauge se centran en estas cantidades que parecen ser constantes de la naturaleza. ¿Por qué los quarks y las otras partículas elementales poseen las masas que tienen? ¿Qué es lo que determina la masa de la partícula de Higgs? ¿Qué es lo que determina la unidad fundamental de carga eléctrica o la intensidad de la fuerza del color? La contestación a esos interrogantes no puede venir de las teorías gauge existentes, sino de una teoría más general.

En la búsqueda de una tal teoría es natural aplicar una vez más una receta que ya ha demostrado su utilidad. Habrá que seguir con el programa de buscar simetrías globales y explotar las consecuencias derivadas de su conversión en simetrías locales. Este principio quizá no sea necesario, pero vale la pena ensayarlo. De la misma forma que la teoría de Maxwell combinó electricidad y magnetismo, y que el modelo de Weinberg-Salam-Ward ligó electromagnetismo y fuerza débil, quizás podría también encontrarse una teoría más amplia que abarcara el modelo de Weinberg-Salam-Ward y la

cromodinámica cuántica. Tal teoría podría construirse, en principio, sobre el modelo de las teorías gauge ya existentes. Es preciso hallar una simetría más general de la naturaleza; al convertirla en una simetría local, aparecerían entonces las fuerzas fuertes, débiles y electromagnéticas. En el proceso pueden aparecer quizás nuevas fuerzas, demasiado débiles para que se hayan observado.

Ya están en curso investigaciones sobre tales teorías. Ultimamente se han concentrado en simetrías que permiten las transformaciones entre quarks y leptones, la clase de partículas que incluye el electrón. A mi entender los esquemas propuestos hasta el presente no resultan convincentes. La gran simetría que ellos presuponen debe romperse para explicar las disparidades observadas entre las fuerzas, y esto exige la presencia de varios campos de Higgs. La teoría resultante comprende tantas constantes arbitrarias de la naturaleza como las teorías menos globales a las que pretende substituir.

Recientemente se ha postulado una forma distinta y más ambiciosa de realizar la unificación; se trata de la llamada "supersimetría" y "supergravedad". Reúne en una única categoría partículas con varias cantidades de momento angular; hasta ahora partículas con spin diferente quedaban incorporadas a categorías distintas. Aunque está por demostrar la utilidad de las teorías supersimétricas, su fecundidad parece muy prometedora. Ofrecen una descripción muy restrictiva de algunos centenares de partículas, gravitón incluido, en términos de sólo unos pocos parámetros ajustables. Hasta la fecha, los resultados no se parecen demasiado al mundo físico conocido, pero eso mismo ocurría en la primera teoría de Yang-Mills en 1954.

La forma de unificación buscada con más ahínco y tenacidad es la reconciliación de las distintas teorías cuánticas de campos con la teoría general de la gravitación. El campo gravitacional parece conducir inevitablemente a teorías cuantificadas que no pueden ser renormalizadas. A distancias extraordinariamente pequeñas (10^{-33} centímetros) y a intervalos de tiempo muy pequeños (10^{-44} segundos), las fluctuaciones cuánticas del mismo espacio-tiempo cobran importancia, y hacen cuestionable el significado del continuo espacio temporal. Ahí están los límites actuales no sólo de las teorías gauge, sino también de todas las teorías físicas conocidas.

Subsistencia en el norte de España durante la última glaciación

En una cueva próxima al golfo de Vizcaya se han descubierto restos que cubren un intervalo de 12.000 años y que reflejan las preferencias alimentarias del hombre en el Pleistoceno

Lawrence Guy Straus, Geoffrey A. Clark, Jesús Altuna y Jesús A. Ortea

Entre los habitantes del oeste de Europa, durante la última glaciación, había grupos de cazadores y recolectores que vivían en la franja costera que bordea el golfo de Vizcaya. Al abrigo de las cuevas donde acamparon, quedan restos de animales que habían cazado o pescado para su alimentación. Una de esas cuevas, La Riera, situada a mitad de camino entre Oviedo y Santander, guarda un registro, de particular riqueza, de lo que fue la dieta humana a lo largo de un intervalo que duró 12.000 años, y cuyo punto de origen puede situarse hace 21.000 años. En el descubrimiento e identificación de esos restos animales, unidos a otras señales de ocupación humana y de las fluctuaciones ambientales de La Riera, han intervenido los firmantes del artículo y otros muchos colegas. En ese trabajo investigador invertimos los últimos cuatro años. Fruto del mismo es la reconstrucción del registro dietético humano a lo largo de esos miles de años del Paleolítico Superior y del Mesolítico, que aquí presentamos. Registro que bien pudiera llevarnos a una cuestión de mayor alcance: ¿cómo pasó el hombre de entonces de la caza y recolección a la agricultura y ganadería?

La cornisa cántabra está formada por las provincias de Guipúzcoa, Vizcaya, Santander y Asturias. Las pruebas de ocupación paleolítica se hacen particularmente densas en una extensión de 350 kilómetros, desde el río Nalón, en la parte media de Asturias, hasta los Pirineos. La franja cántabra mide entre 25 y 60 kilómetros de anchura y la separa de la meseta la cordillera cantábrica. Muchas cumbres de los Picos de Europa alcanzan los 2600 metros de altura, y estuvieron cubiertas por hielo durante el último período glacial. En la época de máxima glaciación, el nivel del mar se hallaba unos 100 metros por debajo de su posición actual; pero debido a que allí la plataforma continental es estrecha y abrupta en su pendiente, la llanura costera no se prolongó, en general, más allá de 10 kilómetros. Comparado con el clima más severo de la cordillera y el de la meseta continental interior, el clima marítimo templado hizo de la costa un lugar atractivo para cazadores y recolectores.

Los granos de polen recogidos en cierto número de yacimientos cántabros prehistóricos indican que la vegetación dominante en la región, en el tiempo de máxima glaciación, era de tipo abierto. Abundaban pastizales y arbustos de tamaño pequeño, como los brezos; aunque había también algunos pinos. En los períodos más benignos aparecieron árboles de hoja caduca, como el abedul, el avellano, el roble y el olmo. La fauna incluía mamíferos adaptados a las condiciones de campo abierto, entre ellos el bisonte europeo, el caballo y algún reno. El ciervo era un animal muy común, incluso se le veía en espacios desprovistos de arbolado, tanto en la llanura costeña como en los amplios valles de montaña, hacia el sur. Parece que la cabra montés y su pariente menos corpulento, el rebeco, prosperaron en las laderas de las colinas rocosas y en las zonas de montaña no cubiertas de hielo. (El ciervo adulto podría pesar entre 200 y 250 kilogramos; la cabra montés adulta no llegaría a la mitad de ese peso.) En la zona rocosa de la costa abundaban los moluscos y los peces de la familia del salmón en ríos y estuarios costeros.

Antes de que los cazadores del Paleolítico Superior hicieran acto de presencia en la región cántabra, habitaban la zona hombres cuyos útiles de piedra pueden clasificarse como musterienses y cuyos restos corresponden al tipo neanderthal, hombres, pues, anatómicamente premodernos. Altuna y Straus analizaron los restos animales de varios yacimientos musterienses de la zona; dedujeron que los cazadores del Paleolítico Medio mataban pocos ciervos y cabras montesas. Preferían cobrarse, en el momento oportuno y según parece, piezas mayores: caballos (peso de un adulto: 350 kilogramos), bóvidos (peso de un adulto: 800 kilogramos o más) y dos especies de rinocerontes (de peso indeterminado). Según se desprende, esos cazadores desconocían técnicas avanzadas de captura, como podrían ser las batidas; y quizá se alimentaron, en una proporción significativa, con métodos carroñeros. En la región cántabra sólo se han descubierto 13 yacimientos musterienses, y todo parece indicar que, en el Paleolítico Medio, la población humana era más bien reducida. Ello explica que los neanderthales no precisaran recurrir, con seguridad, a estrategias complicadas para su propia subsistencia. Los hallazgos arqueológicos del Paleolítico Superior, por contra, inducen a pensar de un modo diametralmente opuesto.

El yacimiento de La Riera está emplazado en la cara sur de una sierra baja que se levanta en la llanura costera, a unos 1,5 kilómetros de la línea de la costa actual, en la Asturias oriental. Se halla cerca de la entrada de un paso montañoso que conduce, a través de la Sierra de Cuera, hasta el valle que corre de este a oeste en la falda de los Picos de Europa. Un eminente prehistoriador español, el Conde de la Vega del Sella, fue quien la descubrió y quien primero se puso a excavar en ella. Desde el punto de vista de su riqueza en registro prehistórico



ENTRADA DE LA CUEVA DE LA RIERA, situada a 100 kilómetros de la ciudad de Oviedo, en el norte de España; está casi oculta por la vegetación. Usada como refugio por grupos de cazadores y recolectores en el Paleolítico Superior y Mesolítico, en ella hay restos de los animales utilizados en su dieta.

UN EXCAVADOR examina, a la luz de una lámpara de gas, el suelo de la cueva de La Riera. Las cuerdas superiores definen la parrilla de control. El yacimiento ha proporcionado muchos artefactos líticos y unos 32.000 dientes y fragmentos de huesos y astas de mamíferos de posible identificación.



LAS COSTAS DEL GOLFO DE VIZCAYA de España y Francia ***(izquierda)*** **difieren notoriamente. La llanura costera española es estrecha y, extensa, la francesa. Los yacimientos franceses, cuyos nombres han permanecido en la nomenclatura de las culturas del Paleolítico Superior y Mesolítico, comprenden a Aurignac y Mas d'Azil en los Pirineos, cerca de la frontera con España, Solutré hacia el nordeste y La Madeleine en la Dordoña. La región cántabra**

constituye un yacimiento de particular valor.

Straus y Clark se responsabilizaron de la dirección de las excavaciones en la propia cueva. Altuna se encargó del análisis de los restos mamíferos y Ortea del examen de los moluscos. Otros 18 colegas contribuyeron en el proyecto desde su respectiva especialidad: análisis de isótopos del oxígeno en conchas y travertinos, análisis por carbono 14, sedimentología, petrología, botánica, examen del polen y similares. Debemos citar, entre ellos, a Henri Laville, de la Universidad de Burdeos, ocupado en el estudio de los sedimentos, Arlette Leroi-Gourhan, del Musée de l'Homme de París, investigadora de los granos de polen, y M. M. de la Hoz, de la Universidad de Oviedo, quien tuvo a su cargo el análisis de los restos de peces.

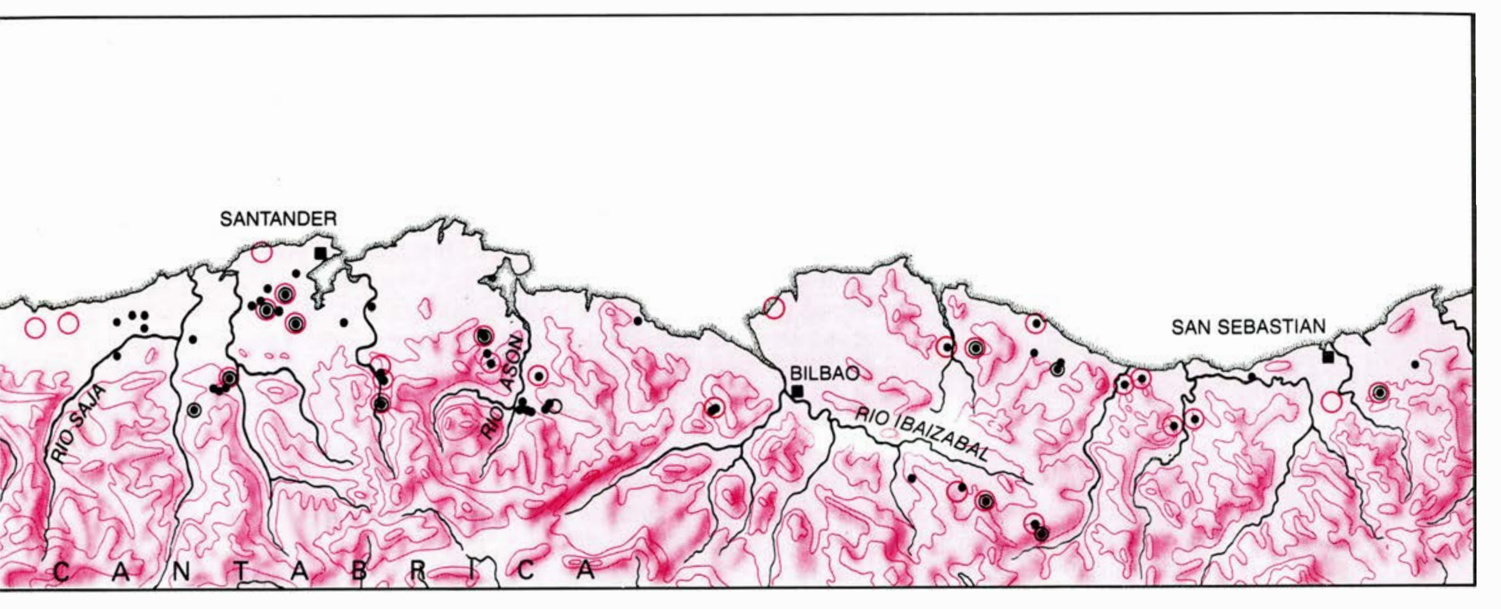
Durante las cuatro campañas de excavación referidas, que se concentraron en una sección intacta del fondo del vestíbulo de la cueva, se dejaron al descubierto 36 niveles distintos en una secuencia vertical de poco más de dos metros y medio. Muchos niveles incluían varias superficies de ocupación, o suelos de habitación, donde yacían útiles de piedra, hueso y asta, además de restos animales y cantos para hogares donde preparar la comida. El inventario de los útiles líticos registra más de 55.000 piezas, de las cuales unas 3000 están talladas. El número de fragmentos de huesos y dientes de mamíferos supera los 200.000, unos 32.000 de los cuales pueden asignarse a especies concretas.

En la parte inferior de nuestra secuencia estratigráfica hay un horizonte que quizá pueda atribuirse, a partir de algunos útiles de piedra y hueso, a una cultura del Paleolítico Superior Temprano: la auriñaciense. En Cueto de la Mina, un yacimiento próximo de la Asturias oriental, es un nivel auriñaciense el que subyace a los estratos solutrenses, cultura ésta perteneciente también al Paleolítico Superior; no obstante, seguimos sin saber la identidad exacta del nivel basal de La Riera. En cualquier caso, será obviamente más antiguo que los niveles inmediatamente superiores, cuya datación se ha fechado en unos 21.000 años antes de la actualidad y de los que les separa un hiato deposicional.

El análisis del polen muestra que el medio ambiente del nivel era relativamente templado y húmedo. La vegetación estaba representada por árboles como el avellano y el roble junto con una gran variedad de herbáceas y arbustos de la familia de las Compuestas. A pesar de esos indicios de un ambiente parcialmente boscoso, los huesos de animales son, sobre todo, de especies de campo abierto: caballos y bóvidos (con buena probabilidad, el bisonte europeo). Abundan también los restos de ciervo.

Los siguientes 16 niveles principales de La Riera contienen útiles que deben adscribirse al solutrense. Los dos primeros niveles solutrenses (correspondientes a los números 2 y 3) se depositaron en condiciones climáticas frías y húmedas; y son pobres en materiales líticos y en restos animales. Predominan, entre los últimos, huesos de caballo y, en menor proporción, huesos de cabras montesas, ciervos y bóvidos.

Los contenidos de los tres niveles siguientes (números 4, 5 y 6) cambian por completo. Sugieren un proceso de desarrollo de técnicas de caza más especializadas, técnicas que acabarían por combinarse con estrategias de subsistencia, complejas y diversificadas ya, del Paleolítico Superior Tardío y del Mesolítico de la región cántabra. Por la época en que se depositaron esos niveles reinaba un ambiente típico de glaciación máxima. El descenso consiguiente del nivel del mar alejó la línea de costa, desde la cueva, unos 10 kilómetros. El clima era muy duro y, salvo unos pocos pinos, la vegetación se reducía a brezos, compuestas y herbáceas, como grupos más destacados. Los tres niveles de la cueva que suministraban esta información son finos y de breve extensión lateral. No aparecen señales características de residencia, verbigracia, hoyos u hogares excavados. Muchos de los instrumentos líticos recogidos allí son únicos en el sentido de estar tallados con variedades de pedernal y cuarcita que no suelen darse en otros niveles de la cueva. También escasean los nódulos y grandes lascas corticales, que constituyen los primeros materiales que se desechan en la manufactura de útiles de piedra, en tanto que abundan rebanadas de desecho pequeñas, que se producen en las etapas finales del trabajo o en el proceso de pulimentado. Ello induce a pensar que muchos de los



entre el río Nalón y la frontera francesa *(derecha)* contiene más de 100 yacimientos del Paleolítico Superior y Mesolítico *(véase la clave, a la izquierda)*. Los yacimientos del Paleolítico Superior Tardío y Mesolítico son los más numerosos y ocupan una gran variedad de nichos ecológicos. La cueva de La Riera, situada al norte de los Picos de Europa, fue descubierta por un eminente prehistoriador español, el Conde de la Vega del Sella.

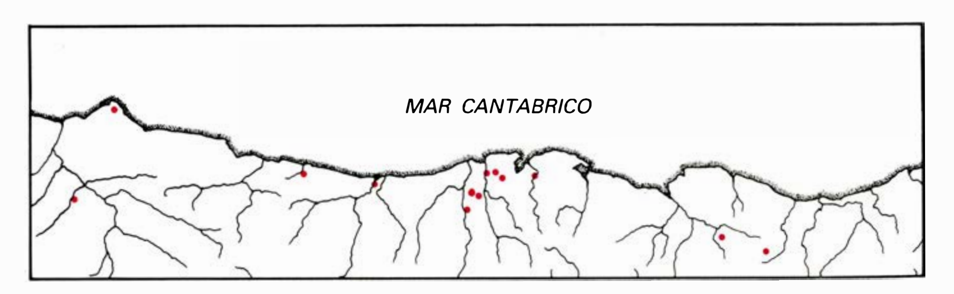
útiles se fabricaron a partir de material semimanufacturado traído hasta La Riera desde otros sitios.

La mayoría de los materiales líticos son puntas en «muesca» pertenecientes a la cultura solutrense. Con excepción de unas cuantas intactas, menudean las de base de muesca. Y hay en cantidades excepcionales para esta fase temprana del Paleolítico Superior hojas minúsculas que han sido «rebajadas» (retocadas en un borde). La escasez de otro tipo de utillaje confiere al muestrario una apariencia particular, como si se hubiera utilizado la cueva para breves incursiones de caza, siendo mucho menor otra clase de actividades en la misma.

Impresión que corroboran los huesos de animales encontrados. Predominan los de cabras montesas: entre los tres horizontes suman por lo menos 28 ejemplares. Los hay juveniles y viejos, pero sobre todo adultos jóvenes. De lo que cabe deducir que los cazadores no mataban las cabras una a una, sino que conducían una manada, «ojeándolas» ladera arriba, hasta ponerlas a tiro de los hombres apostados a su espera. En estos niveles se encuentran relativamente pocos fragmentos de testa animal (partes internas del cuerno, dientes superiores y trozos craneanos o mandibulares), lo cual implica que se despiezaba la presa en el lugar donde le habían dado muerte y allí se abandonaba la cabeza. (Un ejemplo antagónico observamos en Rascaño, yacimiento magdaleniense abierto en un acantilado de la provincia de Santander. Aquí cuernos y cráneos de las cabras montesas están completos, debido quizás a que el campamento de los cazadores se hallaba próximo al lugar de sacrificio.) Además de esos restos copiosos de cabras montesas en dichos niveles descubrimos también los de dos rebecos, por lo menos. La presencia de ambas especies en las laderas rocosas confirma la hipótesis de que los cazadores se servían de La Riera como de campamento base para sus expediciones por las vertientes de la Sierra del Cuera, a kilómetro y medio tierra adentro.

Las cabras montesas y los rebecos no fueron las únicas presas. Los cazadores prosiguieron en su caza de caballos y ciervos, aunque en número menor. Raramente, sin embargo, se observan en esos niveles restos de bóvidos. Al propio tiempo, aparecen ahí, por primera vez y de forma clara, dos grupos marinos comestibles. Se trata, por un lado, de peces de la familia del salmón, que desovan en aguas dulces: el salmón, el reo y la trucha; es el grupo menos destacado. El más importante corresponde a moluscos comestibles, en su mayoría grandes lapas (una variedad de *Patella vulgata*), frecuentes en estuarios y lugares abrigados, que alcanzan, en su crecimiento máximo, una media de cuatro centímetros en su eje mayor. La abundancia de lapas es tal que, en peso total, sobrepasan el 70 por ciento con respecto al peso de los huesos de mamíferos, en el Nivel 4. Otra prueba de que los cazadores de cabras montesas se acercaban a la costa es la existencia en dicho nivel número 4 de dos huesos de foca, pertenecientes al esqueleto de un ejemplar recogido quizás ya muerto. No deja de sorprendernos esa constancia en su explotación de los recursos marinos si nos fijamos en que una correría hasta la costa, desde La Riera, suponía entonces un viaje de unos 20 kilómetros, entre ida y vuelta.

Los dos niveles solutrenses siguientes (correspondientes a los números 7 y 8) aportaron muestras para análisis mediante carbono 14, lo que permitió



YACIMIENTOS DEL PALEOLITICO MEDIO en la región cántabra. Son escasos y no forman un conjunto claro de asentamiento. Los huesos de animales indican que los cazadores sacrificaban y amontonaban en vertederos pocas piezas, pero de gran tamaño. Esta estrategia satisfacía sus necesidades.

su datación entre 21.000 y 20.000 años antes de la actualidad. El clima seguía siendo frío. Predominaba una vegetación de matorral con escaso arbolado en el paisaje. Los dos niveles encerraban un repertorio variado de útiles de pedernal. El Nivel 7 proporcionó asimismo pruebas de una manifiesta técnica en la preparación de la carne a comienzos del solutrense. Apareció un hoyo excavado en la superficie de habitación que había sido reforzado con cantos. Se veían los cantos cuarteados por el fuego y, entre ellos, huesos de animales rotos y carbonizados, pudiéndose distinguir fragmentos de cráneo y asta de ciervos. Ello nos permitió deducir que los cazadores cocinaban al menos algunas de sus piezas cobradas: introducían en el hoyo los cuerpos despiezados, cabeza incluida, y los asaban con cantos calentados al fuego.

Los huesos de animales procedentes de ambos niveles demuestran que la caza de la cabra montés seguía siendo práctica común. Por contra, apenas si aparecen restos de caballo: sólo se han podido identificar siete ejemplares de esta especie. Continúan escaseando huesos de bóvidos y los de rebecos se muestran ahora menos frecuentes. Se cazaron algunos corzos; al parecer, estos animales, mucho menores y más específicamente adaptados a zonas boscosas que los ciervos, no encajaban bien en el paisaje abierto de ese período. Y así, es el ciervo el que predomina en esos niveles: en el número 7 descubrimos restos de 34 individuos y, de 19, en el número 8. La mayoría de los ciervos eran juveniles, pero la cifra de adultos o viejos induce a pensar que, al igual que ocurría con las cabras montesas de los niveles inferiores, los animales fueron ojeados en grupo.

Los cazadores que se resguardaban en la cueva en esa época prosiguieron en sus incursiones costeras para recoger lapas. En el Nivel 7, las cantidades de conchas de *Patella vulgata* resultan impresionantes, aunque en ese horizonte la abundancia de huesos de mamíferos determina que la relación de pesos de concha con huesos sea relativamente baja. En ese nivel apareció también un hueso de foca. La variedad y abundancia de útiles y restos animales que ofrecen los niveles números 7 y 8, sobre todo el primero de los dos, dan fe de una ocupación más intensa, o de una serie de ocupaciones, de la cueva de La Riera durante este período, en comparación con los primeros asentamientos, más especializados y fugaces.

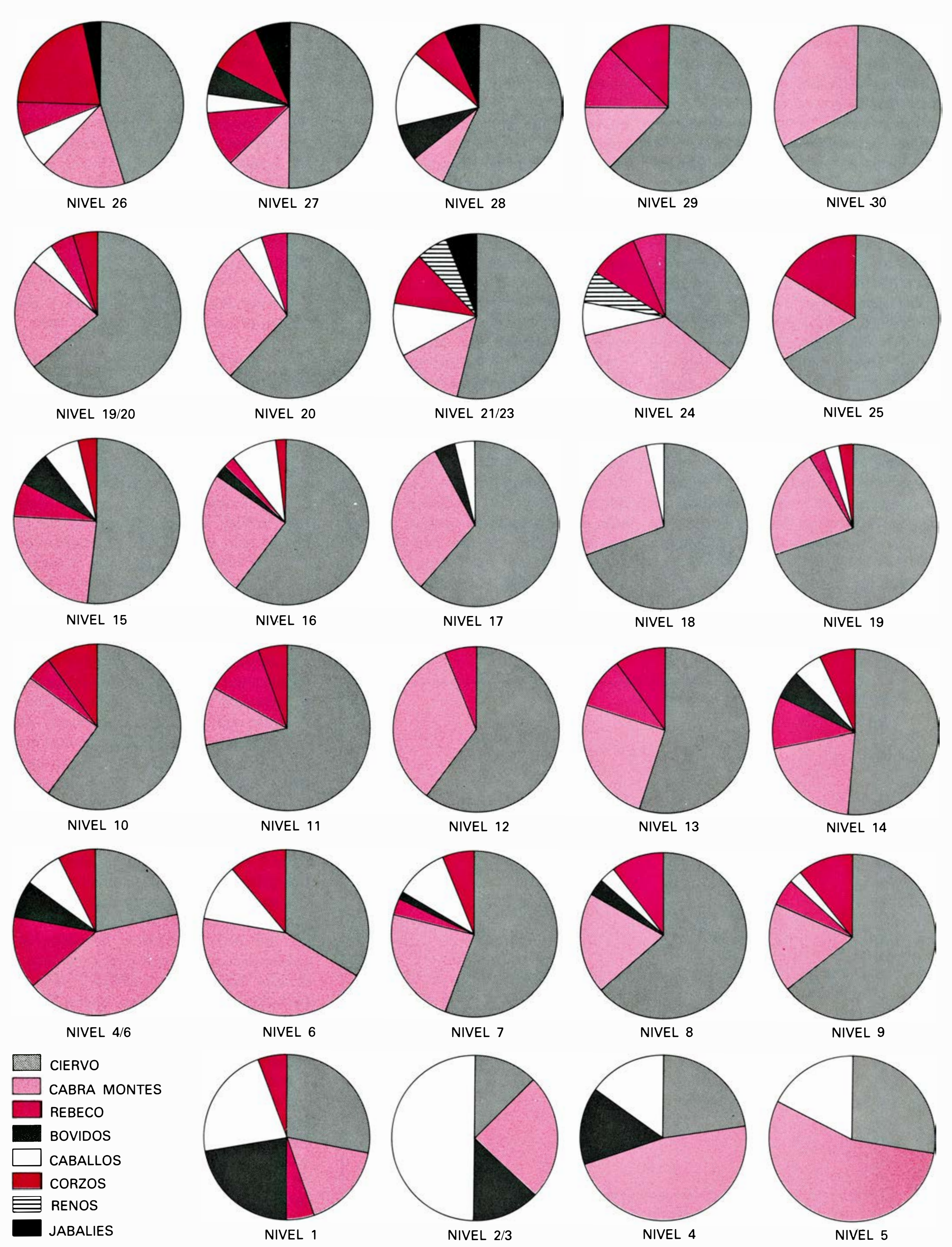
Los siete niveles que siguen (del 9 al 15) corresponden a una extensión de 2000 años del período solutrense. El clima reinante era entonces relativamente templado y húmedo. Crecieron árboles (abedules, avellanos y encinas) hasta formar, al menos, pequeños bosques localizados. La suavidad climática de este intervalo temporal se refleja en la presencia de granos de polen de esos árboles y la deposición y desgaste de sedimentos, que indican un aumento en la temperatura y la humedad. Por si fuera poco, un molusco, el bígaro común *(Littorina littorea)*, que no tolera ambientes relativamente cálidos, desaparece de entre los restos de la cueva.

De estos siete niveles, unos son gruesos y otros delgados; los hay que hablan de breves visitas a La Riera y están los que presuponen un hospedaje más prolongado. Pero todos poseen una característica común, a saber: entre el 80 y 90 por ciento de los restos de mamíferos son ciervos. Esa situación de mayoría se mantiene tanto si se calcula el número mínimo de animales, considerados de forma individual, según muestran los restos, como si atendemos al peso de hueso por especie o simplemente al número total de huesos por especie. El otro 10 a 20 por ciento de restos de mamíferos corresponden, principalmente, a la cabra montés. Los caballos se equiparan ya a los bóvidos en su escasa representación, y a pesar de que aparezcan ya coberturas boscosas, los corzos tampoco menudean.

Salvo en el Nivel 14, los moluscos ofrecen una pobre representación, por más que en esa época el nivel del mar se hallaba probablemente en una cota más elevada y resultaba mucho más fácil el acceso al litoral. Las conchas de moluscos recuperadas en el número 14, un horizonte rico, sobre todo en lo referente a útiles humanos y restos animales,

NIVEL		AÑOS ANTES DEL PRESENTE (A.P.)	POLEN DE ARBOLES (%)	INDICACIONES CLIMATICAS (POLEN, SEDIMENTOS, FAUNA)	
30(T)			48-64	AUMENTA LA TEMPERATURA Y LA HUMEDAD	
29/C	ASTURIENSE	8500 A.P.	12-53		AUSENCIA DE BIGAROS
28	AZILIENSE		(SIN MUESTRAS)		
27		10.500 A.P.	8-22	TEMPLADO, HUMEDO (EPISODIOS FRIOS)	NORDICO RAICES RATAS DE AGUA
26			37		
25			(SIN MUESTRAS)		
24		10.750 A.P.	4,5-16		RENOS
21/23	MAGDALENIENSE	13.000 A.P.	5-11	FRIO, HUMEDO	PINO Y ABEDUL
20			4,8	FRIO, SECO	
HIATO SEDIMENTARIO					
19/20					
19		16.000 A.P.	11	MAS CALIDO, SECO (LOCALMENTE HUMEDO)	ROBLE, OLMO, ABEDUL, AVELLANO, PINO
18			2		
17			3,5		
16		17.500 A.P.	6	FRIO, SECO (DATOS SEDIMENT.)	
15			1,5		
14			2,8		
13			5,5	MAS CALIDO, MAS HUMEDO	ABEDUL, AVELLANO, ENCINA
12			3,3		
11			4		
10	SOLUTRENSE	20.000 A.P.	0,3		AUSENCIA DE BIGAROS
9			4		
8		20.500 A.P.	1,6		
7			3,5	FRIO, SECO (EPISODIOS CALIDOS, HUMEDOS)	NORDICO RAICES RATAS DE AGUA
4/6					
6			0,3		
5			7		
4		21.000 A.P.	4		
2/3			3	FRIO, HUMEDO	
HIATO SEDIMENTARIO					
1	AURIÑACIENSE?	+21.000 A.P.	10,5	TEMPLADO, HUMEDO	

REGISTRO ESTRATIGRAFICO de La Riera, que abarca el intervalo que transcurrió desde hace 21.000 años (unos 19.000 años a. de C.) hasta 8650 años atrás (unos 6650 años a. de C.). Durante 12 milenios, los cazadores-recolectores solutrenses y magdalenienses fueron los principales ocupantes de esa cueva.



RESTOS DE MAMIFEROS TERRESTRES, que se encuentran en cantidades variables en los diferentes niveles de la cueva de La Riera. Esta serie de gráficos circulares muestra el número mínimo de individuos de cada una de las ocho especies de mamíferos presentes en el nivel en cuestión. El número mínimo se expresa en porcentaje del número total de animales presentes ***(véase la clave en la parte inferior izquierda)*****. La mayor parte de la carne del Nivel 1 procedía de caballos y bóvidos. A partir de ese horizonte los ciervos** ***(gris claro)*** **van a predominar casi siempre sobre las cabras montesas** ***(color claro)*****, pero ninguna de las especies restantes, excepto los caballos en los números 2 y 3, superan en constancia a las cabras montesas o los ciervos.**

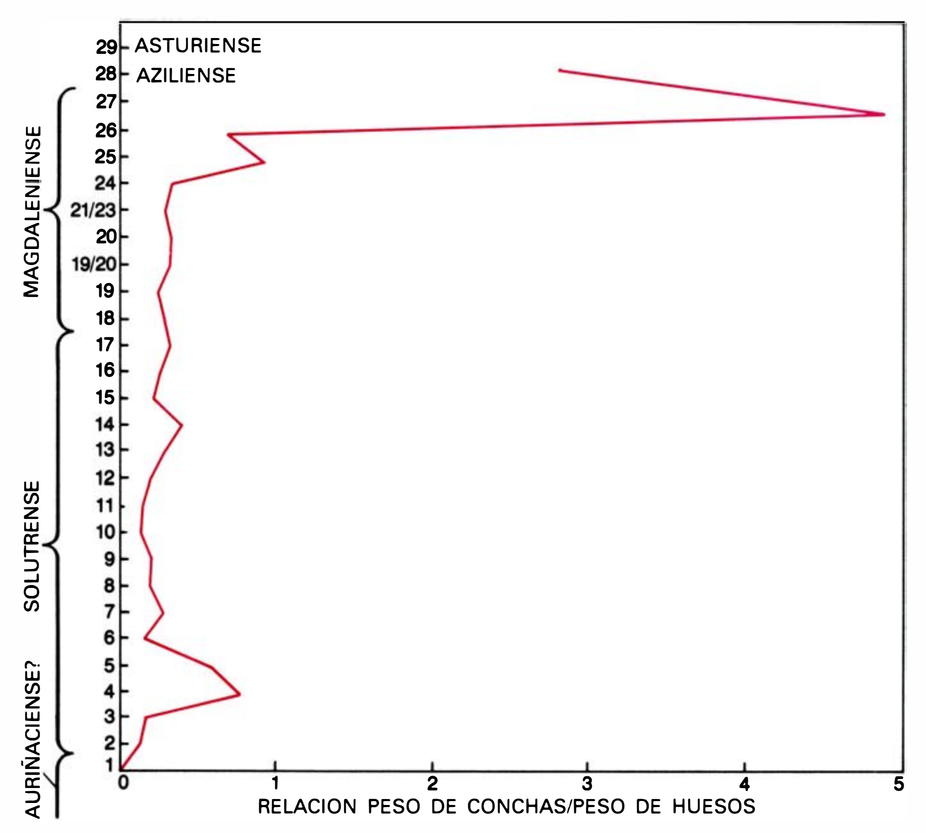
son de aquellas lapas estuarinas de buen tamaño pertenecientes a la especie *Patella vulgata* que cogían los cazadores primitivos. Abundan los cantos cuarteados por el fuego, así como varios hogares esporádicos. Con toda verosimilitud, el asado de la carne constituyó una de las actividades de preparación de los alimentos en que se ocuparon estos cazadores más modernos.

Las herramientas líticas halladas en los siete niveles sugieren que la preparación de alimentos en formas diversas fue realmente una actividad principal en las ocupaciones del lugar. Además de las lascas, presumiblemente para cortar, y los raspadores, que sugieren la preparación del cuero, el inventario de útiles abarca gran número de «denticulados» y lascas melladas, artefactos quizá para ciertos usos en la matanza o para preparar materias vegetales. En esos niveles, las escasas puntas solutrenses son del tipo «hoja de laurel», tal vez puntas de lanza o cuchillos. La disposición diverge diametralmente de lo que observamos en el yacimiento solutrense clásico de Laugerie-Haute, Francia, en donde las hojas de laurel preceden estratigráficamente a la puntas en muesca.

Los dos niveles siguientes (números 16 y 17) encierran materiales pertenecientes a las postrimerías del solutrense de La Riera. El análisis sedimentario permite suponer que el nivel inferior se depositó en un período de clima frío, en tanto que el superior lo hizo en condiciones, de nuevo, relativamente templadas. Ello no obstante, los habitantes continuaron, incluso en el episodio frío, en su caza de ciervos, sobre todo. Así, el Nivel 16 aportó restos de 27 de estos animales y 11 cabras montesas. Podríamos afirmar que no hay más mamíferos en el registro. La propia explotación de los recursos marinos se limitó a la pesca de algunos salmones y reos y de alguna que otra lapa.

Hace unos 17.000 años que se ocupó el Nivel 17. Entre sus artefactos líticos destaca la punta de hoja de sauce solutrense; el 70 por ciento de este útil se encuadra en la misma clase: hojitas de dorso. (En el Nivel 16, las hojitas de dorso totalizan el 24 por ciento de las herramientas.) El predominio de un solo tipo instrumental induce a pensar en un cambio en las actividades de La Riera en ese período o bien en la adopción de una nueva técnica: referida posiblemente a la industria de puntas de flechas.

Los niveles siguientes introducen una nueva cultura: la magdaleniense. Según parece, los niveles del Magdaleniense Inferior (números 18 y 19) se fueron depositando sin solución de continuidad durante un clima relativamente templado, el que dominó a lo largo de los últimos niveles solutrenses. Su datación se remonta 16.000 años. Predominan todavía las hojitas de dorso entre los útiles inventariados y, entre los animales cazados, los ciervos. A éstos les siguen en número las cabras montesas; hay restos vestigiales de corzos, rebecos, caballos y bóvidos. Escasean las conchas de lapas y los únicos restos de peces son de reos.



IMPORTANCIA DE LA ALIMENTACION MARINA en la dieta de La Riera. Se representa en este gráfico según la relación entre el peso bruto de huesos de mamíferos y el de conchas de moluscos. Dicha relación no refleja la aportación real de proteínas de vertebrados e invertebrados, respectivamente. El numeral 1, en el eje de ordenadas, indica una relación de peso de uno a uno. Esta relación se superó de forma considerable durante la fase aziliense. Sin embargo, tanto en el Solutrense Inferior como en el final del Magdaleniense, el peso de conchas alcanza en alguna ocasión el 70 por ciento del peso de restos óseos.

Un extenso hiato separa esos dos primeros niveles magdalenienses del inmediato siguiente: el número 20. Fechado entre 13.000 y 12.000 años antes de la actualidad, ese nivel se formó a lo largo de un episodio de clima frío y seco. Pero los cazadores que se abrigaban en la cueva sacrificaron también más ciervos que cualquier otra especie animal. De un mínimo de 21 mamíferos llegados hasta nosotros por sus restos óseos, los ciervos totalizan 13 individuos; los restantes se reparten así: seis cabras montesas, un rebeco y un caballo.

Aunque la cantidad dietética suministrada por el mar sigue siendo escasa, la variedad es ahora mucho mayor. Además de las lapas estuarinas de cierto tamaño, recogieron *Patellae intermediae*, una especie de proporciones más reducidas, y una especie de caracol, *Monodonta lineata*. Como las lapas más pequeñas, este molusco habita en las zonas del intermareal rocoso de la costa y es raro en estuarios abrigados. Por primera vez se consumieron bígaros en cantidades suficientes para representar una fracción significativa de los moluscos recogidos. También se mariscaron mejillones y erizos de mar, lo que pone de manifiesto que los visitantes magdalenienses de La Riera buscaron su alimento en la línea de costa abierta y en el interior de los estuarios.

Por primera vez, también, se encuentran intactas las conchas de los bígaros y las del caracol *Monodonta*. La explicación es inmediata. Se hirvió o coció abundante copia de los mismos en algún tipo de cacharro y se extrajo la carne de la concha, quizá con algún artefacto punzante, como haríamos hoy.

Del examen del contenido de los tres niveles magdalenienses subsiguientes (correspondientes a los números 21, 22 y 23) se deduce que La Riera sirvió de

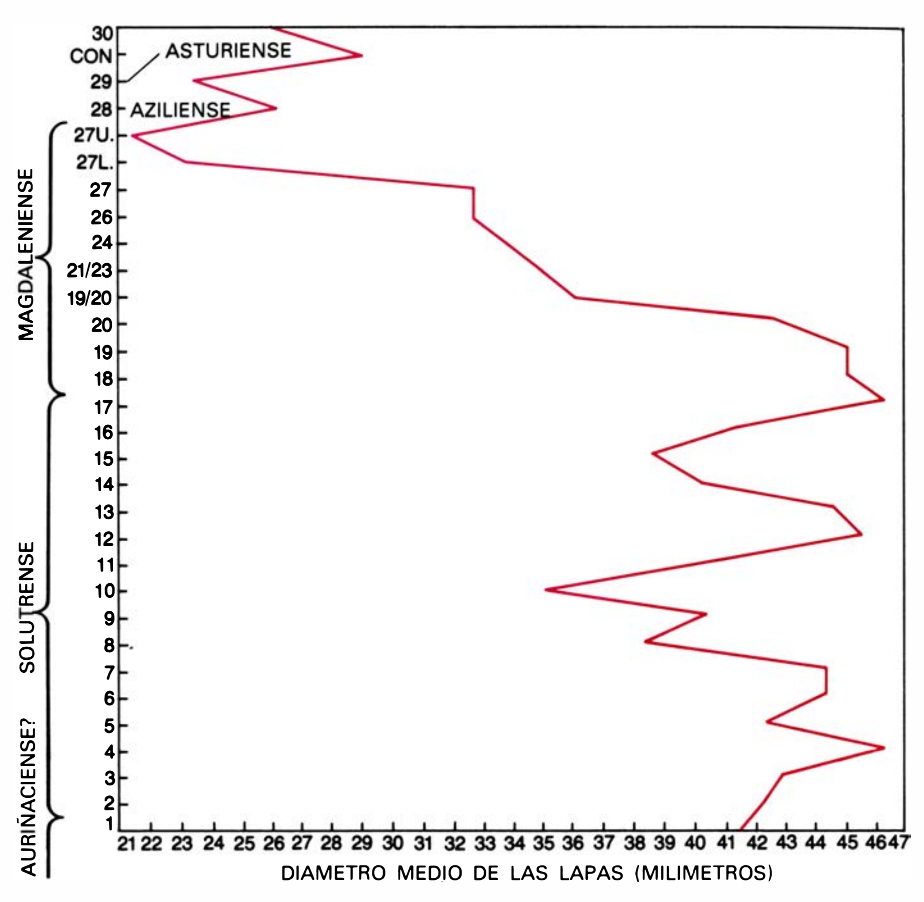
refugio en contadas ocasiones a lo largo de un período que era frío y muy húmedo. Los depósitos son ahora de arcillas de cierto espesor, que permiten inferir que la cueva conoció esporádicas inundaciones de agua. A comienzos de ese intervalo, que duraría en torno al milenio, sufrieron un fuerte gelilevantamiento. El paisaje estaba moteado de parches boscosos, en los que podríamos observar algunos abedules y pinos. Y en ese período inicial había también avellanos, muy pocos; en las postrimerías del lapso abundaba ya el polen del junípero.

Los restos de mamíferos procedentes de estos niveles, limitados en número, muestran una gran variedad. Por primera vez aparecen en La Riera el jabalí y el reno: dos ejemplares de cada uno. No es fácil explicar la coexistencia de huesos de ambas especies. El reno suele vivir en un ambiente frío y abierto, y es raro dar con restos del mismo en esa zona de España. El jabalí prefiere un medio radicalmente opuesto: paisaje boscoso templado y matorral denso. Esa coexistencia manifiesta de ambas especies permite pensar que, en las postrimerías de la última glaciación, la región asturiana presentaba un mosaico complejo de tipos vegetales condicionados por factores variables, como la exposición, altitud y disponibilidad de agua. Aquí también, el ciervo continúa siendo la pieza de caza más cobrada: 16 ejemplares al menos se descubrieron desde el Nivel 21 al 23. Aparecen corzos en número no inferior a tres, idéntica cifra de cabras montesas y cuatro caballos.

Con el Nivel 24 coronamos las fases finales de la ocupación magdaleniense de La Riera. Un arpón de asta, de sección circular, califica al nivel inscrito en el Magdaleniense Superior. El depósito se formó algo después de los 11.000 años antes de la actualidad. Reinaba un clima frío y húmedo. Los pinos constituían la principal especie arbórea y había también manchas de helechos.

Los restos de mamíferos conservados en el Nivel 24 muestran que se cazaban los mismos animales. Los huesos denuncian un número escaso de caballos y, menor, de corzos y rebecos. Los renos hacen su segunda y última aparición en La Riera. Predominan, una vez más, el ciervo y la cabra montés, con once ejemplares por lo menos de cada uno. La mayoría de las cabras eran juveniles.

Se recogieron pocas conchas de moluscos en ese nivel. Por primera vez, sin embargo, aparecen restos de peces marinos, incluidos probablemente los lábridos. Se sigue apreciando la familia del salmón, que a buen seguro se atrapaba con redes en los charcos de marea y en los ríos, como antes. Por contra, los peces marinos habían de pescarse, desde las rocas costeras, con redes y otras artes. Y, por vez primera también, aparecen en abundancia huesos de aves, que son raros en los niveles iniciales de La Riera.



MARISQUEO INTENSIVO DE MOLUSCOS en La Riera. El gráfico muestra una media del diámetro máximo de las dos especies de lapas, cuyas conchas se encuentran en la cueva. Aunque algunas disminuciones de tamaño dan fe de una explotación de especies de lapas más pequeñas, que comienza en los tiempos magdalenienses, otras se deben, probablemente, a un aprovechamiento intensivo del molusco.

A partir de ese Nivel 24, las diferencias culturales tienden a perder sus perfiles característicos.¿Estamos en el Magdaleniense Tardío o en el Aziliense? El Aziliense constituye una cultura de transición que traza un puente tecnológico entre el Paleolítico y el Mesolítico. Por otro lado, los artefactos líticos magdalenienses y azilienses se parecen notablemente. Cualquiera que sea la adscripción por la que nos inclinemos, el Nivel 24 debe separarse del 27 por un intervalo de unos 250 años, e intercalarse, entre ambos, los Niveles 25 y 26.

Aunque el clima seguía siendo relativamente frío, los bosques no cejaron en su expansión. Los análisis del polen revelan que, durante el tiempo de sedimentación del número 27, hará de ello unos 10.600 años, el roble, el olmo y el nogal convivían con el pino, el abedul y el avellano. Proliferaban, de modo especial, los helechos y las Compuestas.

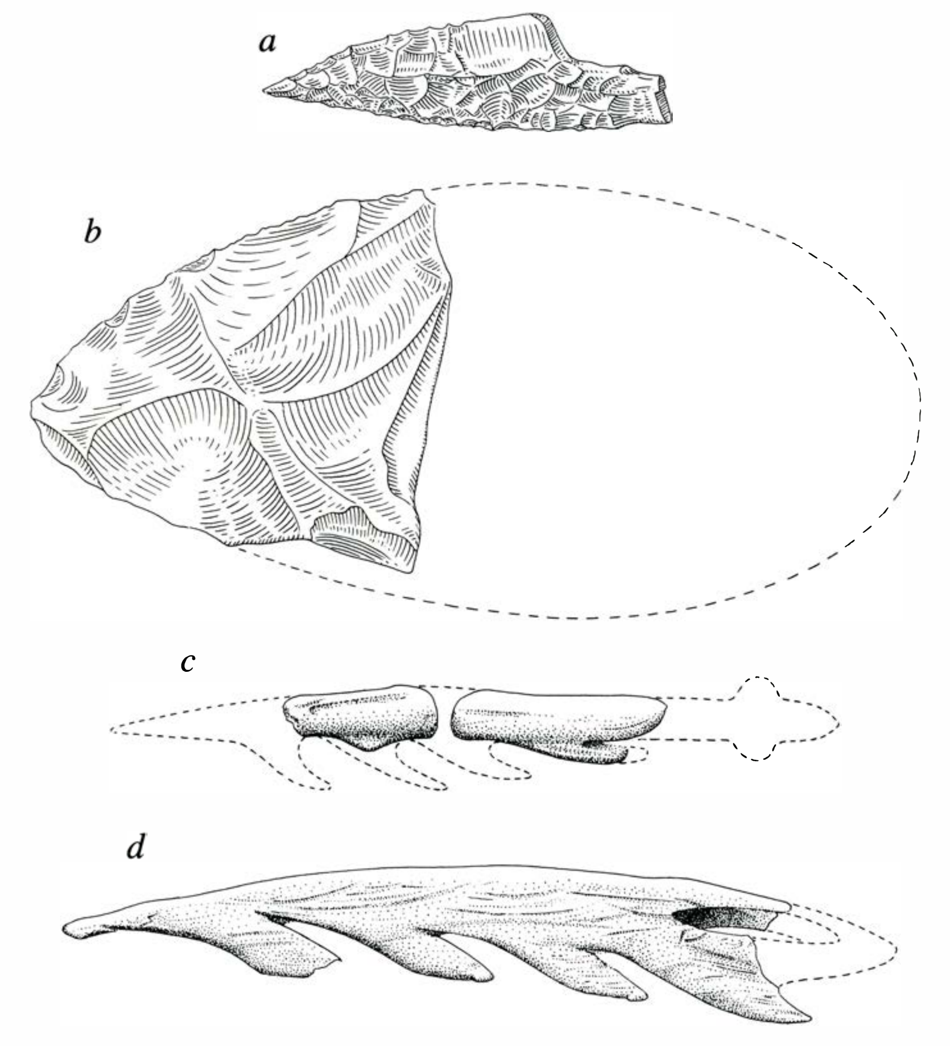
Los pormenores de la dieta que pueden deducirse a partir de los restos animales en esos niveles transicionales resultan bastante claros. Así, en el número 25, donde escasean los huesos de animales, domina el ciervo. De los mamíferos representados en el Nivel 26, los ciervos repiten predominio, aunque su mayoría esta vez no es absoluta: hay trece ciervos, seis corzos, cinco cabras montesas, dos rebecos, dos caballos y un jabalí. En el Nivel 27 la proporción de los ciervos se ha incrementado: suman 28 del total de 56 animales. Straus descubrió en su análisis de la dentadura que había una cantidad destacada de cervatillos. Siguen en orden de abundancia las cabras montesas, los rebecos y los corzos y, más distantes, caballos, jabalíes y bóvidos. Estos últimos constituyen los primeros bóvidos que aparecen desde el Solutrense Superior (Nivel 17). Los números 26 y 27 ofrecen también una cifra respetable de huesos de aves.

En ese período, el consumo de los frutos del mar evidencia algunas variaciones significativas. En las mismas fases iniciales, todavía dentro del Nivel 24, se advierte un descenso en el tamaño corporal de las lapas recogidas con más insistencia, *Patella vulgata*, debido probablemente a una explotación intensiva. Esa tenaz extracción del Pleistoceno Tardío pudo motivar que se mariscaran lapas en sectores más expuestos de la costa, donde no se dieran las condiciones óptimas para un crecimiento máximo. Las lapas sufren todavía hoy una explotación intensiva, que explica su tamaño reducido. Allí donde la contaminación impide su marisqueo indiscriminado alcanzan fácilmente su tamaño máximo, como lo harían las de la última glaciación, a pesar de que hoy la temperatura del agua es más templada.

En los Niveles 25 y 26 hay constancia de una explotación intensiva y continuada de moluscos: el peso de sus conchas aumenta sustancialmente en relación con el peso de huesos de mamíferos. La tendencia se acelera; en el número 27 el peso de las conchas supera al de restos óseos en una proporción de más de 1,3 a 1; de nuevo decrece el tamaño medio de *Patella vulgata*. También las lapas más pequeñas de *P. intermedia* y los caracoles *Monodonta lineata*, que no se aprovecharon hasta el Magdaleniense Medio (Nivel 20), contribuyen ahora de forma significativa al volumen total de moluscos. Se diría que el número 27 constituye un vertedero de conchas. En este nivel hacen su primer acto de presencia los caparazones de cangrejos, y abundan los esqueletos de peces marinos.

Si el arpón de asta de sección circular, hallado en el Nivel 24, definía a este depósito como perteneciente al Magdaleniense Superior, otro arpón de asta, ahora con una sección típicamente plana, caracteriza al Nivel 28 de La Riera como propio de la cultura aziliense. El inventario de restos de mamíferos registra aquí las mismas especies descubiertas en el número 27, con la única salvedad de la ausencia de caballos. Más importantes aún, prosigue, de forma acentuada, el consumo de moluscos; el peso de las conchas sobrepasa al de restos óseos en una proporción de más de 4,6 a 1. Las conchas de las especies de lapas pequeñas abundan tanto como las de especies grandes; por primera vez, hay más caracoles *Monodonta* que bígaros. Hay también más espinas de peces marinos que en oportunidades anteriores.

Nos encontramos a unos 10.000 años de distancia. Se ha cerrado el último período glacial del Pleistoceno. La cueva de La Riera se halla rodeada por bosques que florecen en un clima cálido y húmedo, típico del Holoceno Temprano. Ciervos y corzos, cabras montesas, rebecos y caballos siguen considerándose posibles piezas a cobrar, como se desprende de nuestro limitado muestreo del Nivel 29 y de una sección aparte del «conchero», fechado en el 8650 antes de la actualidad, literalmente cementado con el techo de la cueva. En estos depósitos, relacionados con una cultura mesolítica regional denominada asturiense, los restos de mamíferos son escasos, en tanto que proliferan las conchas de moluscos pertenecientes ya a varias especies. Los bígaros, que no toleran aguas cálidas, han desaparecido de los concheros, pero las restantes especies de moluscos llegan ahora a nueve e incluyen caracoles terrestres. De los dos tipos de lapa, la especie mayor aparece de forma escasa y acaba por imponerse *Patella intermedia*. Se encuentran espículas y caparazones de erizos de mar, caparazones de cangrejos y, entre los esqueletos de peces marinos, los de lenguado, especie que habita en el fondo. Y así, aunque el principal elemento de la dieta era todavía seguramente la carne de venado, y pudieran aprovecharse alimentos vegetales como cereales y avellanas, nuestro registro del menú humano a lo largo de más de 10.000 años termina con una clara preferencia por los frutos del mar.



ENTRE LOS INDICADORES DE TIEMPO DE LA RIERA se incluyen los cuatro tipos de utensilios ilustrados aquí. Las puntas de proyectil *(a)*, con la base mellada, se encontraron en los niveles del Solutrense Temprano, mientras que los niveles del Solutrense Tardío proporcionan puntas en forma de hoja *(b)*. Los arpones de asta con sección circular *(c)* son típicos del Magdaleniense Tardío, en tanto que el arpón de sección plana *(d)* es característica de la cultura postpaleolítica de Mas d'Azil (Francia).

¿Por qué después de cientos de milenios de subsistencia a base de caza y marisqueo adoptó el hombre, sólo en fecha muy reciente, la estrategia alternativa agropecuaria? La razón de ese cambio espectacular sigue siendo objeto de debate. Algunos prehistoriadores relacionan el cambio de estrategia con los cambios ambientales registrados a finales del Pleistoceno; para otros, tiene que ver el aumento de población y la consiguiente presión demográfica de entonces. Lo que quizá

pase por alto es que los habitantes del Paleolítico hubieron de padecer, por mucho tiempo, cambios climáticos y ambientales que afectaron a sus estrategias de subsistencia y demostraron ser capaces de adaptarse a esos cambios según las circunstancias.

La secuencia cultural de la región cántabra proporciona un ejemplo excelente de varios cambios de esos y su repercusión en la subsistencia. Los escasos yacimientos musterienses conservan un registro de dependencia oportunista, con un número reducido de grandes mamíferos. El muestrario del Paleolítico Superior de La Riera y otros yacimientos cántabros, a su vez, comienza cuando no se estimaban todavía los recursos marinos y finaliza cuando esos recursos se explotaban intensamente. En un intervalo de varios milenios, aumentaron otras fuentes alimentarias, cuyo aprovechamiento requirió estrategias especializadas que se apoyaban en la elección de asentamientos que permitieran sistemas complejos de colonización-subsistencia, en las postrimerías del Paleolítico Superior. Muestra de ello es la presencia de abundantes restos de cabras montesas en el Solutrense Temprano de La Riera (Niveles 4 al 6), lo que da a entender que el refugio sirvió de base para la caza especializada de esta especie en su cercano hábitat roqueño.

El registro de La Riera y otros puntos de la cornisa cántabra no apoya la tesis de que los factores ambientales constituyan la causa determinante del cambio en la estrategia de subsistencia. El clima, cálido o frío, húmedo o seco, y el medio, boscoso o despejado, aportaron, qué duda cabe, dos piezas de caza principales a lo largo de los últimos milenios del Pleistoceno: la cabra montés y el ciervo. Los animales de caza de mayores proporciones (caballo y bóvidos) integraron el menú obligado de la época musteriense. Aunque proseguía su caza en número menor, en el Paleolítico Superior, estos animales representaron sólo una parte de la elaborada estrategia de alimentación básica; la dieta reflejaba una caza intensiva de cabras montesas y ciervos y la explotación de moluscos, peces, aves y vegetales. Posiblemente, caballos y bóvidos desempeñaron un papel destacado en el Paleolítico Medio, pues las escasas partidas de cazadores-carroñeros de esa época no necesitaron inventar las eficientes técnicas de la captura masiva, que hubiera requerido la caza de manadas de ciervos y cabras.

Los cazadores del Paleolítico Superior manifiestan una tendencia ascendente a extender su campo de recursos, a pesar de las variaciones ambientales. La tendencia se hace evidente desde el momento en que se comienza a aprovechar las lapas y otros moluscos, a principios del Solutrense. Si el clima reinante era frío, y la costa se hallaba distante de la cueva, o si era cálido, y la orilla quedaba más cercana, no parece, en cualquiera de los dos casos, que la longitud de la correría determinara el que se mariscase o la cantidad de la explotación. Quizás influyó más la división del trabajo, las necesidades y la planificación estacional que las consideraciones del tiempo. A modo de prueba, digamos que algunos de los períodos en que se mariscaron moluscos con más intensidad coincide con episodios fríos, cuando dicho trabajo suponía un esfuerzo mayor.

La única ocasión en que el cambio ambiental condicionó directamente la estrategia de los cazadores parece producirse a finales del Pleistoceno, cuando un aumento de los bosques incidió en un aumento correlativo en la cacería de corzos y jabalíes. ¿Cómo explicar entonces el notable énfasis puesto en el aprovechamiento de frutos del mar en una época más o menos coincidente?

Entendemos que uno y otro incremento participaron de un acontecimiento final en el marco de un extenso registro de estrategias de subsistencia intensificadas y diversificadas, que comenzaron a aplicarse en la llanura costera cántabra en el Solutrense, 10.000 antes de que se cerrara el Pleistoceno. El desarrollo de esas estrategias se vio acompañado de un perfeccionamiento técnico específico de los útiles de caza y marisqueo: redes y trampas, arpones, lanzas arrojadizas y, ya en el Mesolítico, arcos y flechas. Tal copioso registro de la explotación intensiva de recursos no puede correlacionarse con los cambios ambientales. Por otro lado, si el rápido aumento de la lista de yacimientos arqueológicos de la región cántabra, correspondientes a finales del Paleolítico Superior, constituye un dato significativo del aumento de la densidad demográfica, habrá que suponer que la presión de población forzó la utilización intensa de los recursos marinos y la explotación extensiva de todo tipo de fuentes alimentarias, hecho que es evidente en La Riera y a lo largo de la cornisa cántabra. Esas mismas presiones demográficas condujeron, al parecer, a la adopción de métodos agrícolas y ganaderos en la región cántabra, al igual que en el resto de Europa, unos cinco mil años después del Pleistoceno.

La hipótesis del gas de las profundidades terrestres

Hay muchos indicios de que los terremotos liberan gases procedentes de zonas profundas del manto terrestre. Entre estos gases puede contarse metano de origen no biológico, que podría constituir una inmensa reserva de combustible

Thomas Gold y Steven Soter

Hay una creencia generalizada de que las reservas terrestres de hidrocarburos combustibles se consumirán en un futuro cercano, futuro que, para los combustibles más solicitados (petróleo y gas natural), se cifra en unas cuantas décadas y para la hulla, en unos cuantos siglos. Diversos indicios nos inducen a creer que en las profundidades terrestres hay volúmenes enormes de gas natural y que, si pudieran explotarse, constituirían una fuente de hidrocarburo combustible que podría durar millares de años. La hipótesis de que en el interior de la tierra hay mucho gas también proporciona una base unificada para explicar muchos otros fenómenos, que anuncian el desencadenamiento de terremotos o los acompañan, y que fuera de este contexto teórico resultan desconcertantes.

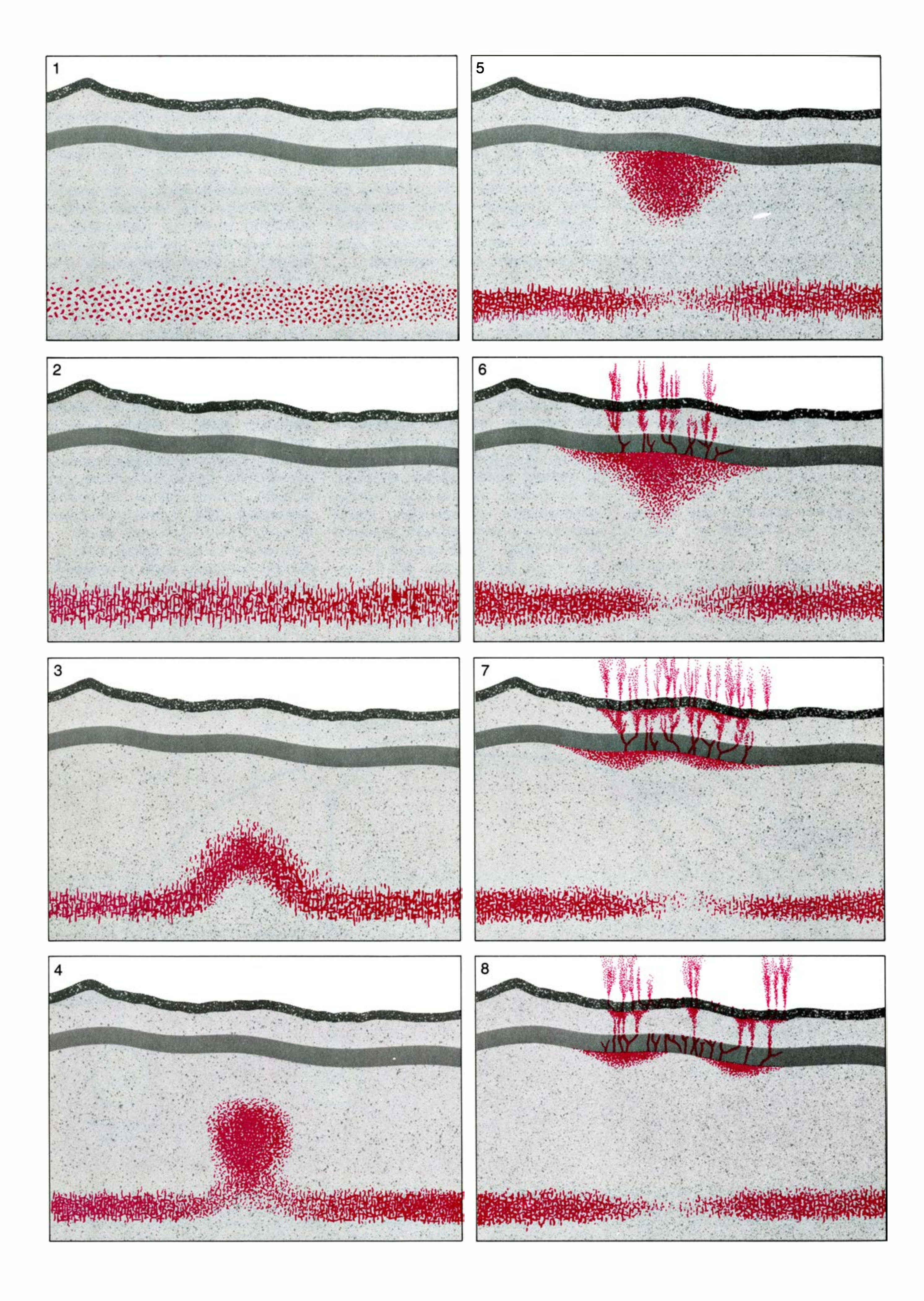
No se conoce la composición exacta del gas, pues los datos empíricos son escasos y su interpretación resulta difícil. Las erupciones volcánicas arrojan a la superficie gases procedentes del interior de la tierra; el agua y el anhídrido carbónico constituyen la mayor parte del volumen de las emisiones, pero, a menudo, se detecta en ellas la presencia de monóxido de carbono, metano, amoníaco, hidrógeno molecular, ácido sulfhídrico y otros gases. Sin embargo, en base a las observaciones realizadas no es posible deducir la composición inicial del gas cuando se encontraba dentro de la tierra, ya que, por una parte, una proporción no determinada de gas volcánico puede estar constituida por volátiles reciclados a partir de sedimentos corticales, y no por gas juvenil que llega por vez primera a la superficie, procedente del manto; por otra parte, gases reducidos o ricos en hidrógeno se habrán oxidado casi completamente en el magma líquido al ascender hacia la superficie y, en tercer lugar, la mayoría de las muestras de gas se han tenido que obtener necesariamente de volcanes que se encontraban en una fase relativamente inactiva, con lo cual, desde el punto de vista químico, las muestras pueden ser poco representativas de los grandes volúmenes de gas emitidos en erupciones explosivas.

Es muy probable que los gases liberados en el transcurso de terremotos sean muestras más fidedignas de lo que hay en la parte profunda de la corteza y en el manto superior. La obtención de muestras de estos gases acaba de empezar, y los datos con los que se cuenta aún no permiten llegar a conclusiones terminantes. Cabe suponer que la composición de los gases de las profundidades varía de un lugar a otro, puesto que la ubicación de los yacimientos minerales en la corteza sugiere que el manto subyacente es muy heterogéneo. Por varios motivos, creemos que uno de los gases principales del interior de la tierra es metano de origen no biológico; en torno a él se centrará la exposición del presente artículo, si bien ello no supone minimizar la posible importancia de otros gases de las profundidades terrestres en los fenómenos asociados a los terremotos.

La idea de la existencia de metano no biológico choca contra la concepción dominante en la geología del petróleo, según la cual casi todo el petróleo y todo el gas natural terrestres tienen un origen biológico. De acuerdo con esta concepción, todo el carbono de los hidrocarburos combustibles procedió originariamente de anhídrido carbónico atmosférico, y la energía necesaria para disociar el carbono y el oxígeno provino de la luz solar, a través de la fotosíntesis realizada por plantas verdes. Algunos de estos compuestos orgánicos habrían quedado enterrados antes de que pudieran experimentar oxidación y se habrían convertido en los materiales fuente del petróleo y del gas. No cabe la menor duda de que este proceso contribuyó a la génesis de gran parte del petróleo que se ha explotado, pero quizás haya algo que añadir a lo contado.

La hipótesis de que la tierra contiene mucho hidrocarburo no biogénico parte de la observación de que los hidrocarburos constituyen la mayoría entre las moléculas provistas de carbono del sistema solar. En su mayor parte, el universo está constituido por hidrógeno, y las pruebas aportadas por el estudio de la cosmoquímica inducen a creer que originariamente la tierra y el resto del sistema solar se condensaron a partir de una nebulosa saturada de hidrógeno. Teniendo en cuenta que los meteoritos proporcionan las claves más fidedignas para conocer la composición original del interior de los planetas, es

HIPOTETICO ASCENSO DE GAS desde las profundidades terrestres, representado de manera esquemática en la página opuesta. El gas, liberado a partir de la provisión originaria de hidrocarburos de la tierra, se difunde lentamente hacia los poros (*1*) que pueden considerarse como oquedades parecidas a burbujas comprendidas entre granos de la roca. Los poros se distienden y quedan interconectados (*2*), creando un "dominio poroso" que empieza a ascender (*3,4*) en virtud de la inestabilidad de la presión. Cuando el dominio poroso llega a una roca más fría y más dura (*banda oscura en 5*) su movimiento se paraliza (*6*). El gas facilita la fracturación de la roca en roturas netas, y parte del gas sale a la superficie. A través de las fracturas originadas por un terremoto (*7*) pueden escapar hacia la atmósfera grandes volúmenes de gas. Después, la lenta difusión del gas (*8*) determina la existencia de una zona progresivamente más extensa, propensa a experimentar sacudidas posteriores. El escape de gas en las tres últimas etapas explica la aparición de muchos fenómenos precursores de terremotos o asociados a ellos.



1
2
3
4
5
6
7
8

significativo el hecho de que la mayor parte de su carbono se encuentre en forma de hidrocarburos complejos que presentan algunas analogías químicas con los alquitranes del petróleo.

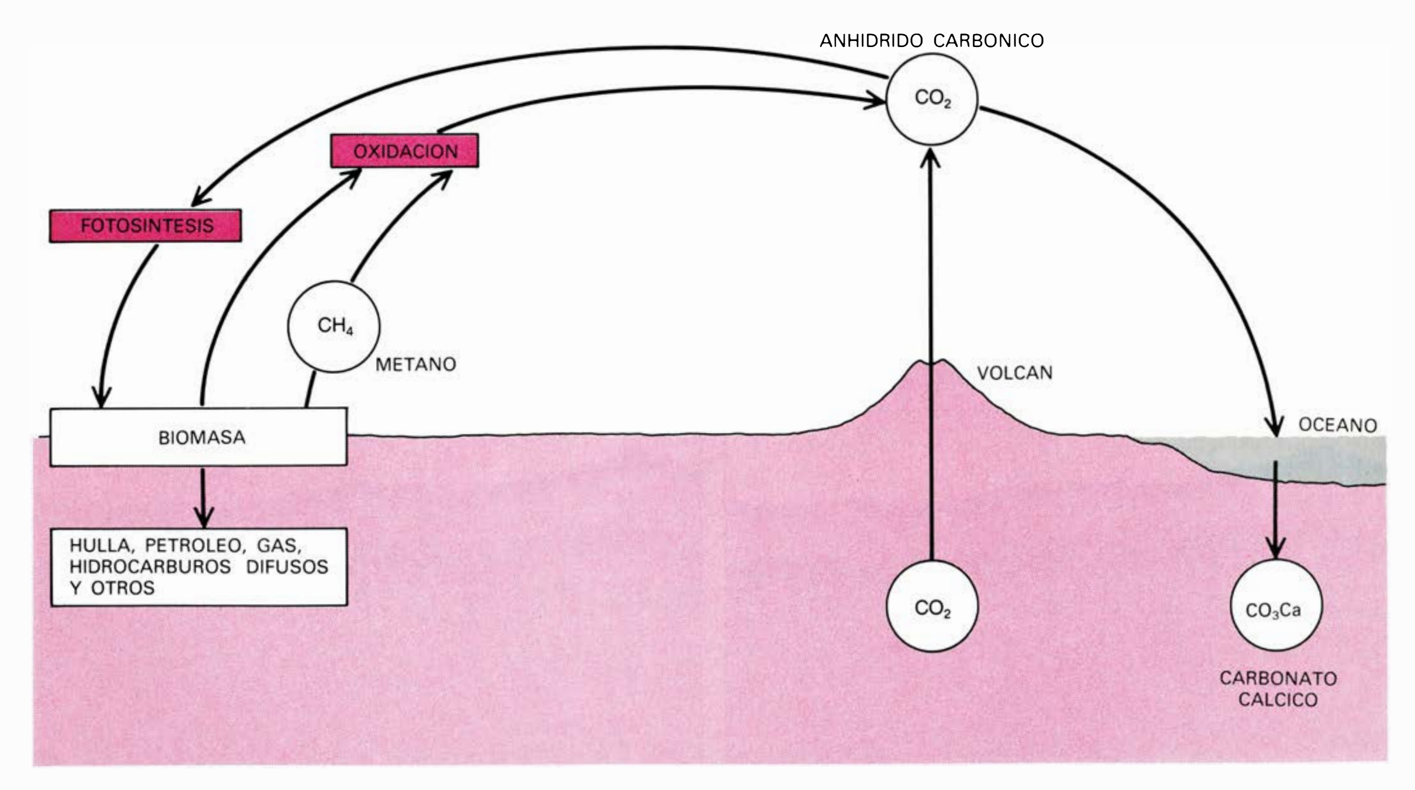
Parece verosímil que la tierra adquiriera una cantidad importante de su carbono por vía de ese tipo de hidrocarburos. Probablemente, la mayor parte del carbono contenido en la primitiva atmósfera terrestre se encontraba en forma de metano (CH_4). Se cree que en las etapas más tempranas del desarrollo de la vida sobre la tierra fue necesaria una atmósfera de estas características. Gracias a la producción subsiguiente de oxígeno libre mediante la fotosíntesis, la atmósfera fue adquiriendo este elemento gradualmente, hasta alcanzar su actual composición, rica en oxígeno. Esta riqueza en oxígeno es la que permite que los hidrocarburos combustibles puedan constituir hoy una fuente útil de energía química, ya que el oxígeno, que es uno de los componentes necesarios para la combustión, está difundido por toda la atmósfera.

¿Qué ocurrió con la provisión de hidrocarburos con que originariamente contaba la tierra? Sugerimos la siguiente hipótesis. Enterrados y sujetos a condiciones de presión y temperatura elevadas, los hidrocarburos habrían liberado metano como principal componente móvil. Este gas, a menudo asociado a gases procedentes de otros materiales fuente, tiende a ascender a la supeficie, en especial a lo largo de zonas de debilidad de la corteza, dejando tras de sí la masa de hidrocarburos pesados. Allí donde en su recorrido el metano atraviesa la lava volcánica caliente, el oxígeno del agua y de algunos óxidos de las rocas lo oxidarán, convirtiéndolo en anhídrido carbónico antes de que penetre en la atmósfera.

Si en su recorrido el gas experimenta una reducción de la presión en una región más fría no fundida, como, por ejemplo, a lo largo de un plano de falla frío, el gas puede llegar a la superficie en su estado reducido originario. (En la atmósfera, podrá mantenerse reducido sólo unos años, hasta que finalmente se oxide y se convierta en anhídrido carbónico.) A lo largo de otros trayectos el metano puede quedar atrapado, durante cierto tiempo, bajo estratos relativamente impermeables, donde contribuirá a los conocidos depósitos de gas natural. Finalmente, el metano que circule por trayectos que lo conduzcan a depósitos de hidrocarburos, incluidos los de petróleo biógeno, se disolverá en ellos. Si permanece allí mucho tiempo, cosa probable, seguramente experimentará cambios químicos, entre ellos, algunos que provocarán la polimerización del carbono y el hidrógeno en las moléculas ya existentes de hidrocarburos.

La mayor parte del carbono contenido en el metano que emigra hacia arriba llegará a penetrar en la atmósfera, bien sea directamente como metano, bien oxidado en forma de anhídrido carbónico. La atmósfera pierde anhídrido carbónico principalmente por disolución y precipitación en los océanos. Las rocas sedimentarias de la corteza terrestre contienen un enorme volumen de carbono, la mayor proporción del cual se encuentra en forma de caliza (carbonato cálcico, CO_3Ca). El carbono abunda mucho más en los sedimentos que en las rocas ígneas, de las que derivan los sedimentos. Este "exceso" de carbono debe de haberse trasladado desde el interior hasta la superficie bajo las formas correspondientes a los principales gases estables del carbono (anhídrido carbónico y metano), si bien aún no es posible determinar cuáles podrían haber sido las proporciones de ambos gases.

Si todo el carbono sedimentario reducido se originó a partir de anhídrido carbónico degasificado, deberíamos hallar el correspondiente exceso de oxígeno tanto en los sedimentos como en la atmósfera. Pero este oxígeno sobrante no aparece. La provisión de carbono en forma de hidrocarburo



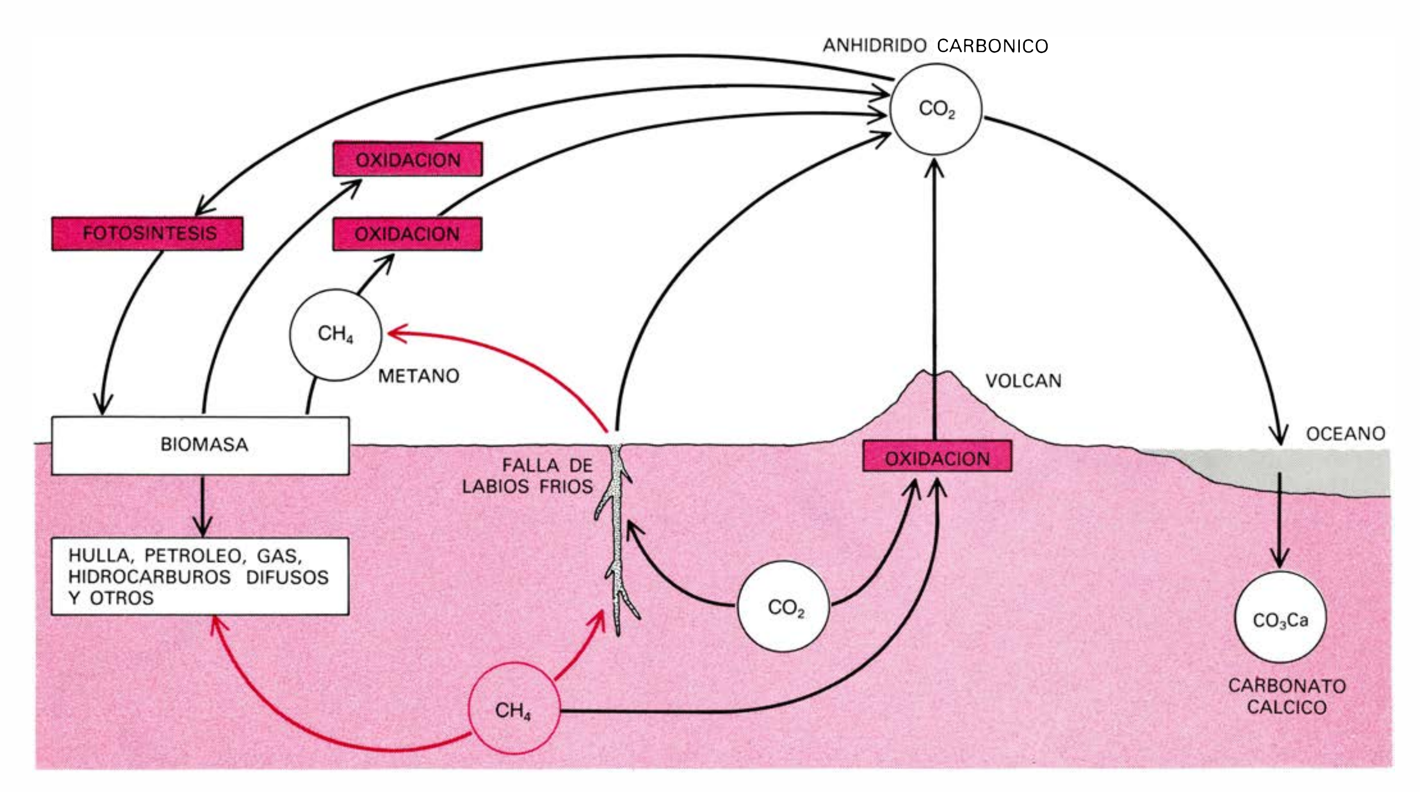
PROVISION DE CARBONO de la tierra, valorada según la explicación convencional (*izquierda*) y según la hipótesis de que a gran profundidad, en el interior de la tierra, hubiera metano que contribuyera, por una parte, a alimentar la atmósfera y, por otra, a incrementar y mantener el volumen de yacimientos de gas natural, hulla y petróleo (*derecha*). Según la concepción convencional, todos estos yacimientos tienen un origen biológico, y el carbono de los hidrocarburos combustibles procede del anhídrido carbónico de la atmósfera. Las plantas habrían disociado el carbono y el oxígeno mediante el

evitaría el problema. En realidad, la degasificación del metano puede ser un importante proceso global que todavía esté actuando. Si el volumen que queda en el interior es comparable al volumen que debe haber ascendido para proporcionar todo el carbono superficial o una parte substancial del mismo, la cantidad de metano existente todavía a gran profundidad debe ser enorme, en comparación con la que representan todos los yacimientos biológicos de carbono. Sin duda, los depósitos principales se encuentran a una profundidad excesiva para que puedan ser perforados mediante sondeos; ahora bien, aunque sólo se considere el metano que queda temporalmente atrapado a niveles accesibles en su trayecto de ascenso, lo más probable es que su volumen sea muy grande. Por consiguiente, importa identificar los trayectos a lo largo de los cuales el gas llega a niveles accesibles.

A la idea de que en zonas profundas de la tierra pueda haber hidrocarburos suele objetarse que estos materiales se oxidarían fácilmente, pasando a anhídrido carbónico, en virtud de las elevadas temperaturas que allí reinan. Pero este argumento soslaya dos consideraciones importantes. En primer lugar, la enorme presión de confinamiento, muy grande incluso a profundidades moderadas, favorece la estabilidad del metano y dificulta la oxidación. En segundo lugar, aún en el caso de que las rocas contengan oxígeno en una forma capaz de oxidar el metano, el oxígeno estará disponible en cantidades significativas sólo en rocas líquidas. Si los trayectos de las continuas corrientes de gases se limitan a fisuras de las rocas sólidas, las superficies accesibles quedarán desprovistas de oxígeno muy pronto, y la mayor parte del gas seguirá manteniéndose no oxidado.

Otro argumento que suele esgrimirse es el de que la presencia de porfirinas y de otros residuos moleculares de organismos vivientes en muchas formaciones petrolíferas prueba que todo el petróleo procedió de la descomposición de sedimentos orgánicos. Pero son muchas las rocas sedimentarias ricas en residuos biológicos. Y si a una de estas rocas llega petróleo no biológico procedente de abajo y éste permanece en ella impregnándola durante unos cuantos millones de años a temperaturas y presiones elevadas, seguramente el petróleo quedará contaminado con substancias biológicas derivadas de los sedimentos.

El químico británico Sir Robert Robinson ha escrito que "no se puede afirmar con excesivo énfasis que el petróleo no presente una composición correspondiente a la que sería de esperar de una modificación de productos biógenos, y todos los argumentos basados en los constituyentes de petróleos antiguos encajan igualmente bien, o quizá mejor, con el concepto de una mezcla de hidrocarburos primitivos a los que se hayan agregado productos biógenos". En realidad, no creemos que sea concluyente ninguna de las pruebas que suelen alegarse para defender un origen exclusivamente biológico del petróleo. El cuadro que propugnamos es el de un origen doble: una parte de los hidrocarburos procedería de sedimentos orgánicos enterrados, y otra parte, probablemente mucho mayor, se habría agregado a aquellos hidrocarburos mediante el acceso de corrientes de metano no biológico. Por lo general, el metano se considera como químicamente no reactivo, pero bien podría ser que lograra polimerizarse en forma de petróleo bruto bajo condiciones adecuadas de temperatura, presión y acción catalítica (incluida, quizá, la acción microbiana). Si esto es así, entonces cabe pensar que una corriente ascendente puede agregar lentamente metano a un depósito preexistente de material biológico e incrementar su volumen hasta convertirlo en un yacimiento de petróleo comercialmente valioso. Este proceso no tendría por qué ser muy eficaz; incluso si la mayor parte del gas no queda atrapado sino que escapa a la superficie, la



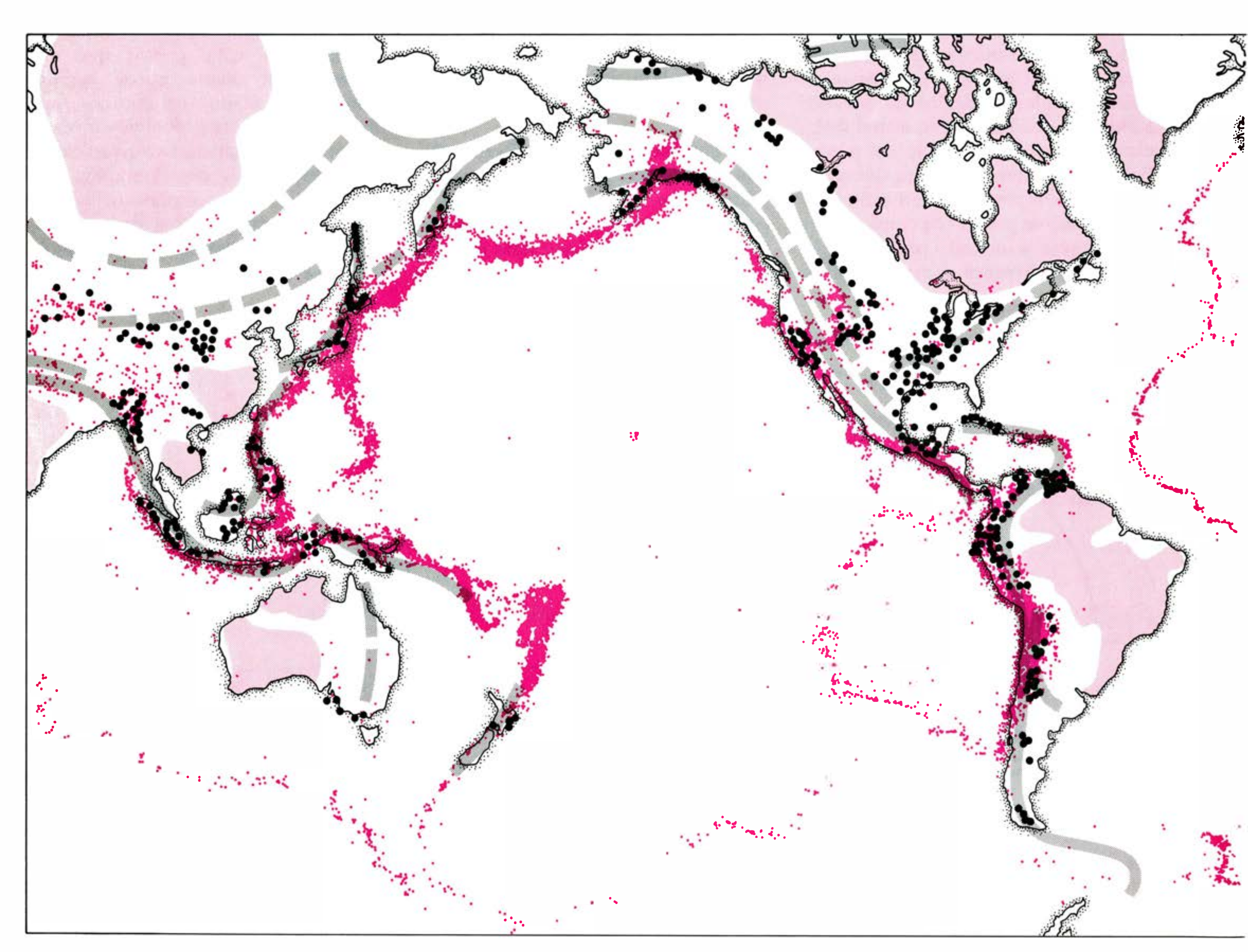
proceso de la fotosíntesis, y la parte de estas plantas que quedó enterrada antes de oxidarse habría formado depósitos de hidrocarburos. En la segunda hipótesis se supone que, además de la reducción de anhídrido carbónico por fotosíntesis, una fuente de metano no biógeno, situada a gran profundidad, contribuye directamente a la formación de depósitos de hidrocarburos. Parte de este metano es oxidado y transformado en anhídrido carbónico al llegar a la superficie a través de lava volcánica caliente a baja presión. El metano puede penetrar en la atmósfera atravesando fallas que separan rocas frías.

persistencia en el transcurso de períodos geológicos de tiempo de un flujo siquiera moderado permitiría un enriquecimiento suficiente para provocar la acumulación que suponen las vastas masas de hidrocarburos halladas en los yacimientos de importancia comercial.

El incremento por vía química del volumen de los yacimientos de hidrocarburos tendría una repercusión positiva de realimentación, puesto que cuanto más creciera el depósito, tanto más aumentarían las probabilidades de que quedara atrapada una molécula de metano ascendente. Este mecanismo podría ayudar a explicar por qué los escasos grandes campos petrolíferos que se conocen son tan vastos, comparados con todos los demás. De los varios millares de campos petrolíferos comerciales, 33 (25 de ellos situados en el Oriente Medio) contienen aproximadamente la mitad de todo el petróleo bruto mundial explotable conocido.

Examinemos ahora algunas de las pruebas que atestiguan el escape de metano desde el interior de la tierra. Los lugares que al parecer vale la pena analizar son las fallas y fisuras corticales de los bordes de placa, según la concepción de la tectónica de placas; estas fallas y fisuras deben constituir las mejores vías de comunicación entre el interior y la superficie terrestre. Y en verdad parece que la posición de los hidrocarburos está claramente relacionada con la de estos lugares. Mediciones efectuadas en aguas que cubren bordes de placa y zonas de valles de fractura (rifts) indican la existencia en ellas de grandes concentraciones de metano disuelto. Por ejemplo, el contenido en metano de las salmueras que ocupan la parte más profunda del mar Rojo es unas 1000 veces mayor que el del agua marina normal. Los fluidos hidrotermales que ascienden de grietas abiertas en el fondo marino de la dorsal del Pacífico oriental contienen elevadas concentraciones de metano. El lago Kivu, situado en el valle de fractura continental de Africa oriental, contiene unos 50 millones de toneladas de metano disuelto, cuya existencia no puede explicarse por una fuente microbiana adecuada, con la que no cuenta. Nos parece probable que todas estas aguas estén recibiendo metano no biológico que se difunde a través de fisuras corticales profundas.

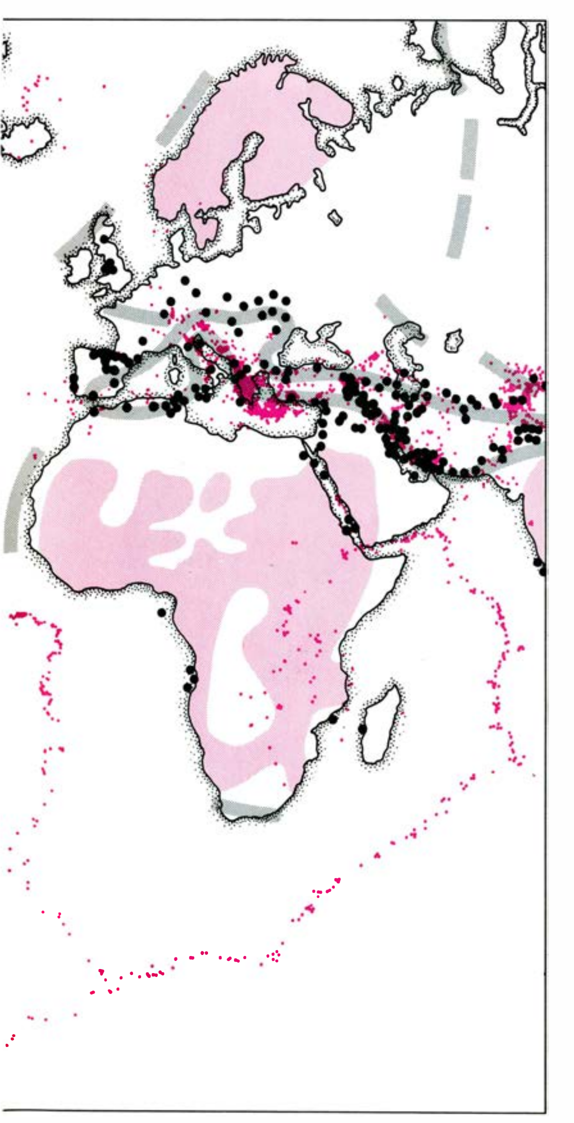
Otra línea argumentativa que conecta los hidrocarburos no biológicos con estos accidentes tectónicos es la sorprendente correlación entre las principales regiones petrolíferas y ricas en gas y las principales zonas de actividad sísmica del pasado y actual. Muchos de los campos petrolíferos se encuentran a



CORRELACION entre las principales regiones productoras de petróleo y gas y las zonas de actividad sísmica, actual o del pasado. En el mapa puede verse la posición de los grandes escudos o masas rocosas (*color claro*), de los principales cinturones de deformación (*gris*), de terremotos importantes (*puntos de colores*) y de diversos rezumaderos de petróleo (*puntos negros*). Muchos de los yacimientos de hidrocarburos conocidos (a saber: Alaska, Texas, Caribe, México, Venezuela, golfo Pérsico, Urales, Siberia y sudeste asiático) se encuentran localizados en cinturones de deformación. La

lo largo de líneas de fallas activas o antiguas. La mayoría de los rezumaderos naturales de petróleo y gas se hallan ubicados en regiones sísmicamente activas. Los rezumaderos de gas más espectaculares son los "volcanes de barro", que son colinas (algunas, verdaderas montañas) cuyos sedimentos están abovedados por las erupciones intermitentes, y en ocasiones violentas, de gas, que a veces es anhídrido carbónico, pero en la mayoría de los casos se trata de metano casi puro. Casi todos los volcanes de barro se localizan en las líneas de grandes fallas activas o en sus inmediaciones, y en algunos lugares son paralelos a alineaciones de volcanes reales.

Cabe suponer que las regiones sísmicamente activas pueden experimentar varios miles de seísmos importantes en el transcurso de unos cuantos millones de años. Se podría pensar que en una de estas regiones la reiterada fracturación de las rocas liberaría, o casi vaciaría, cualquier acumulación de petróleo y de gas, y que ello ocurriría en períodos de tiempo cortos, comparados con la edad de los estratos que la aprisionan. En cambio, la realidad muestra que los campos de petróleo y de gas presentan una asociación patente con estas regiones propensas a los terremotos. La asociación induce a pensar que las fallas que se extienden hasta gran profundidad pueden proporcionar un conducto adecuado para la provisión continua de corrientes de metano no biológico ascendente. Es más, la migración hacia arriba de metano y de otros gases en zonas de falla puede contribuir al desencadenamiento de terremotos.



correlación concuerda con la hipótesis de que desde grandes profundidades ascienden hidrocarburos de origen no biológico por fisuras de la corteza, los cuales se sumarían a los biógenos.

Un terremoto se produce a consecuencia de la liberación del esfuerzo a que está sometida una roca superficial cuando ésta experimenta una fracturación súbita; ello conlleva una propagación rápida de una grieta y un deslizamiento brusco entre los dos bloques separados por la grieta. (Se cree que los esfuerzos se acumulan lentamente, pero todavía no se han identificado claramente las fuerzas motrices.) Hace ya tiempo que los sismólogos señalaron una dificultad con la que se tropieza cuando se pretende explicar los terremotos de foco profundo. A más de cinco kilómetros de profundidad, la presión que ejerce la carga representada por la cobertera rocosa es tan grande que resulta imposible que una grieta se abra por sí misma. Bajo estas presiones, si el esfuerzo de cizalla es excesivo, la roca no se quiebra súbitamente por fracturación, sino que simplemente se deforma. A profundidades mayores, el aumento de la temperatura reduce aún más la capacidad de la roca para desarrollar grietas. El esfuerzo se liberaría a través de una fluencia plástica continua y no se podría producir fractura. Y, sin embargo, se ha registrado la existencia de terremotos cuyo foco se situó a profundidades de hasta 700 kilómetros, y su "rúbrica" sismográfica ha demostrado que aquéllos, al igual que los de foco somero, conllevan una fracturación discontinua súbita y un deslizamiento.

La presencia de gas a gran profundidad podría resolver esta contradicción. Si las rocas profundas poseen poros intercomunicados que se mantienen abiertos por el gas a la presión ambiental, los poros pueden constituir la base de una grieta, con tal de que ésta se forme bajo una presión del gas suficiente como para mantenerla abierta y permitir el deslizamiento de ambos labios. El gas a presión barre eficazmente la roca y ello posibilita el deslizamiento continuo, como ocurre cuando la presión de confinamiento es baja. La grieta puede entonces propagarse, liberando la energía de deformación y produciendo un terremoto.

Si el terremoto tiene una intensidad suficiente para fracturar el subsuelo y llegar a la superficie, el mismo gas contará con una vía de escape y originará algunos de los fenómenos peculiares detectados que han acompañado a muchos de los grandes terremotos. Algunos de ellos son llamas que salen del suelo, "faros sísmicos", intenso burbujeo en masas acuáticas, aire sulfuroso, ruidos explosivos y silbantes, ensordecedores, y ondas visibles que se desplazan lentamente a lo largo de terrenos aluviales. Hay centenares de relatos independientes, recopilados en el transcurso de los siglos, que dan fe de la existencia de estos fenómenos, observados en muchas partes del mundo.

La existencia de las llamas indica que en muchas ocasiones el gas que emana durante un terremoto es combustible; lo más probable es que se trate de metano o de hidrógeno. Según los reportajes sobre el terremoto del valle de Owens que azotó California en 1872, "inmediatamente después de la gran sacudida, personas cuya veracidad está por encima de toda duda... vieron mantos de fuego sobre las faldas rocosas de las montañas Inyo a unas millas de distancia. Estas llamas, observadas en varios lugares, oscilaban adelante y atrás, con manifiesta independencia del suelo, como vastas antorchas; se mantuvieron sólo unos minutos". La realidad de las llamas se verifica por las secuelas físicas que suelen dejar. Por ejemplo, durante el terremoto de Sonora (México) en 1887, varias personas afirmaron haber visto llamas y luego se hallaron ramas quemadas por encima de las líneas de fisura del suelo.

El gas se autoincendiaría con chispas originadas por adquisición de carga electrostática por parte de granos de polvo arrastrados en la corriente, como ocurre en la ignición espontánea de gas en la erupción de volcanes de barro. La formación de llamas puede explicar también algunos de los faros sísmicos descritos en las sacudidas sísmicas nocturnas, faros cuya apariencia se debe a que son observados desde una distancia de algunos kilómetros. Otros fenómenos análogos, tales como destellos, bolas de fuego y luminosidad difusa en

el cielo, pueden deberse a efectos electrostáticos ligados a la descarga súbita de grandes volúmenes de gas en la atmósfera.

También puede justificarse el enérgico burbujeo del mar y de otras masas acuáticas observado en el transcurso de grandes terremotos. Por ejemplo, durante el gran terremoto chileno de 1960, observadores que se encontraban en distintos puntos de la costa, a lo largo de un tramo de 450 kilómetros, informaron que el mar parecía estar hirviendo. Sería de esperar que se produjera un espectáculo como éste en caso de que, en un terremoto, se liberaran grandes cantidades de gas.

El metano es inodoro, pero el ácido sulfhídrico es un constituyente común del gas natural y la presencia de éste podría explicar el olor sulfuroso que se siente después de un terremoto. Dado que el ácido sulfhídrico es soluble en agua y muy tóxico para los peces, la erupción de gas natural cargado de sulfhídrico puede también explicar algunos de los casos en los que se observó la presencia de peces muertos flotando en el agua después de grandes maremotos.

Respecto a la propagación de ondas superficiales lentas por un suelo aluvial, los casos registrados coinciden con muchos terremotos importantes. La descripción es siempre la misma: ondas como las olas del mar, por lo general de varios decímetros de altura, barren distancias cuya longitud es muchas veces mayor que la de la onda. Creemos que estas ondas lentas de gran amplitud sólo pueden propagarse si la cobertera aluvial está levantada con respecto a la roca dura subyacente; es decir, el fenómeno sería debido al empuje ejercido por el gas sometido a presiones de centenares o millares de atmósferas, que manaría a través de las grietas originadas en la roca dura en el momento del terremoto. Este gas puede levantar fácilmente la cobertera aluvial, menos frágil que la roca, y por consiguiente no se agrieta tan fácilmente. Una vez levantada, la cobertera puede propagar ondas de gravedad lentas.

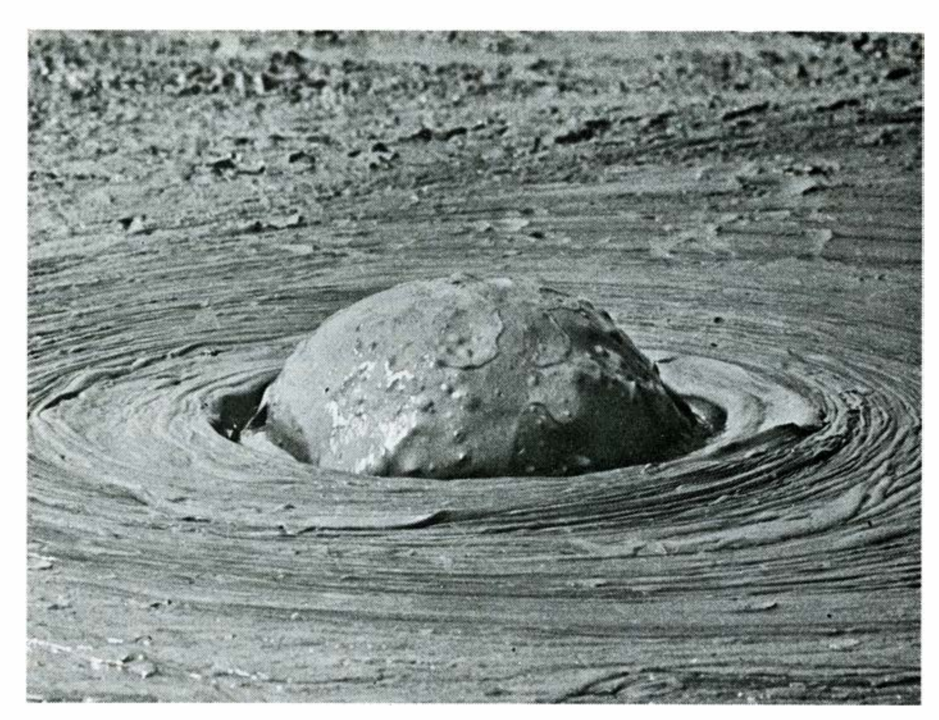
Los tsunamis (inmensas olas marinas provocadas por maremotos que suelen ser terriblemente destructoras) pueden constituir un fenómeno análogo. En general se considera que los tsunamis se originan por un desplazamiento súbito de una enorme extensión del fondo marino a lo largo de una distancia vertical comparable a la altura de la ola. Bajo este supuesto habría que concluir que una masa rocosa, en ocasiones de más de 100 kilómetros de longitud, ha tenido que ser levantada (o hundida) una altura de un metro. Es de suponer que esta masa se extiende hasta la región focal del terremoto (situada a 10 kilómetros o más de profundidad en la mayoría de los casos), lo que implica que el trabajo gravitacional necesario para efectuar este movimiento tendría que ser decenas de millares de veces superior a la energía encerrada en la ola marina. En



ESTE VOLCAN DE BARRO, que se encuentra cerca de Baku, URSS, está asociado a erupciones de gas, lo mismo que otros situados en sus cercanías o en muchas otras partes del mundo. Cheyildag, que así se llama, se mantuvo inactivo durante 50 años, hasta que en 1970 explotó a través de la cobertera rocosa. A la explosión le siguió la aparición de una gran llamarada. Al mismo tiempo empezó a salir de la tierra una masa de arcilla que formó el montículo que aquí vemos (de unos 70 metros de diámetro). Los volcanes de barro se forman principalmente por erupciones de gas y, en general, no van asociados a temperaturas elevadas; suelen estar en las líneas de grandes fallas activas. Las llamaradas de un volcán de barro alcanzan alturas de hasta dos km.

cambio, la energía total adscrita al terremoto, a juzgar por la energía sísmica radiada, se cifra sólo en unas 10 a 100 veces la de la ola. No es posible adscribir al terremoto un contenido energético mucho mayor sin superar, con notable diferencia, la energía máxima de deformación elástica que puedan haber almacenado las rocas.

Al parecer, necesitamos recurrir a un mecanismo diferente para poder explicar los grandes tsunamis. Una erupción de gas en el fondo oceánico y su ascenso a la superficie crearían un movimiento ondulatorio del tipo requerido. Por otra parte, ello requeriría cantidades de energía mucho menores. Una energía sísmica sólo diez veces mayor que la energía del tsumani parece, pues, posible.



ZONA ACTIVA de uno de los volcanes de barro de las proximidades de Baku, representada por la burbuja visible en esta foto. Algunos de los volcanes de barro de los aledaños de Baku miden centenares de metros de altura y varios kilómetros de diámetro. Las erupciones importantes de los volcanes de barro suelen coincidir con terremotos. La mayoría de ellas están constituidas principalmente por metano.

Hasta ahora no se ha demostrado que ninguno de los efectos mencionados haya sido provocado por erupciones de gas durante terremotos, pero, al menos en los fenómenos de llamas y de burbujeo del agua, es difícil imaginar una alternativa verosímil. Según la concepción convencional, puede esperarse que durante los grandes terremotos se produzcan erupciones ocasionales de gas, porque la trituración del subsuelo rocoso podría liberar cualquier bolsa local de gas confinado procedente de los estratos adyacentes. Sin embargo, pueden citarse pruebas que demuestran que el papel que desempeña el gas en el desencadenamiento de un terremoto es algo más que el de un factor pasivo. En realidad, creemos que la mayoría de los fenómenos precursores de terremotos ya conocidos pueden explicarse mejor en función de un aumento de la presión ejercida por el gas de profundidad con bastante antelación al desencadenamiento de un terremoto.

Muchos de los fenómenos precursores sólo se detectan mediante instrumentos. Entre ellos se cuentan cambios en la velocidad de las ondas sísmicas al propagarse por el subsuelo, en la inclinación y altitud de la superficie, en la composición química de los gases del suelo y del agua subterránea y en la emanación del gas radón. El tiempo comprendido entre la aparición de un fenómeno precursor y el desencadenamiento del terremoto varía entre unos minutos y varios años.

Por lo general, el análisis de estos precursores se plantea en términos de dilatancia de las rocas, es decir, la abertura en las rocas de microgrietas cuando el esfuerzo de cizalla se aproxima al valor crítico que determina la fracturación. Convenimos en que un aumento de la porosidad de la roca y una expansión concomitante de su volumen contribuirían a la aparición de la mayor parte de los precursores. Sin embargo, la dificultad con que tropieza este tipo de interpretaciones estriba en que, a profundidades de decenas de kilómetros, donde se originan la mayoría de los terremotos, la roca no podría por sí misma ofrecer la resistencia suficiente para vencer la presión ejercida por la carga rocosa suprayacente y mantener abiertas las grietas. Sin la presencia de un fluido a alta presión que mantuviera abiertas las grietas, la dilatancia no podría ni siquiera iniciarse.

Sugerimos que esto es precisamente lo que hace el gas de profundidad. Nuestro modelo de fracturación a gran presión ambiental presupone el cumplimiento de dos requisitos para que se desencadene un terremoto de foco profundo: (1) un esfuerzo de cizalla suficiente para producir fracturación si la misma roca se encontrara a una profundidad somera y (2) un gas a alta presión, capaz de mantener abiertas las grietas incipientes y, por consiguiente, de contrarrestar el enorme rozamiento de la carga suprayacente. Es la invasión por parte de un gas de estas características procedente de abajo lo que transforma la roca dúctil en frágil, provocando simultáneamente su expansión y los fenómenos de dilatancia asociados.

Cuando el gas de profundidad invade una región que más tarde él mismo convertirá en susceptible de sufrir un terremoto, parte del gas puede trasladarse a la superficie, perturbando el agua subterránea y alterando la conductividad eléctrica del suelo y la composición de los gases del suelo. En particular, el efecto precursor de emanación de radón parece exigir la presencia de un gas vehiculador que circule por el subsuelo. El radón es un gas traza, muy escaso, producido principalmente por la desintegración radiactiva del uranio contenido en las rocas. Su propia vida media radiactiva se cifra en sólo 3,8 días, por lo cual, antes de desintegrarse, sólo podría difundirse por sí mismo, a través de los poros del suelo, unos cuantos metros. En contraste con ello, antes de que se desencadenaran algunos terremotos se detectaron incrementos substanciales en el volumen de radón emitido en superficie, incluso a distancias de 100 kilómetros del epicentro. La explicación más sencilla es la de que el radón no es más que un indicador conveniente de un gas mucho más abundante que lo arrastra consigo, gas que, de no ser por el radón, pasaría desapercibido, pero cuyo flujo es suficiente para atravesar las profundidades en las que se hallan las fuentes de radón y aflorar en superficie en cuestión de días.

Existen otros signos precursores de terremotos, además de los que sólo pueden detectarse mediante instrumentos. Algunos son tan obvios y patentes a los sentidos que se han

identificado desde la antigüedad. Creemos que también estos efectos los produce el incremento de flujo de gas a través del subsuelo. Algunos de estos precursores "macroscópicos" son los ruidos explosivos sordos de origen desconocido, el comportamiento extraño de los animales, los aumentos anómalos locales de temperatura, los humos sulfurosos (a veces acompañados de una niebla peculiar), el burbujeo y otras perturbaciones del agua en los pozos y las llamas sobre el suelo.

Desde luego, se podría argumentar que estos fenómenos precursores relacionados con el gas se producen porque el gas se desprende de los poros expansionados de rocas someras sujetas a una deformación creciente antes de que acontezca el terremoto. Sin duda, una parte del gas podría atribuirse a esta causa, pero dicha fracción del gas difícilmente puede explicar que las manifestaciones precursoras características sean súbitas e irregulares y se presenten en épocas en las que los aparatos medidores de la deformación más sensibles no revelan nada excepcional. Por otra parte, a menudo los precursores son patentes a mucha distancia de la región epicentral de deformación máxima.

Puesto que el gas responsable de los precursores observados suele aparecer antes de que se aprecie ningún cambio en la deformación, aquél debe representar una provisión nueva cada vez, y no una mera redistribución local en respuesta a una deformación que experimenta cambios. Además, los efectos precursores aparecen simultáneamente en extensas regiones. Consideradas conjuntamente, estas observaciones inducen a pensar que el gas procede de una fuente sita a gran profundidad.

El extraño comportamiento que muestran los animales antes de que sobrevenga un terremoto, comportamiento sobre el cual existe una abundante documentación, también puede interpretarse en términos de una emanación de gas. Muchos animales tienen un agudo sentido del olfato que les permite detectar los olores "terrestres" resultantes del ascenso de gases a través del suelo. Otros animales pueden oír ruidos de baja frecuencia originados por el gas, y los animales de morada subterránea pueden responder a la amenaza de asfixia cuando su hábitat es invadido por aire enrarecido.

Si tanto para la dilatación como para el movimiento subsiguiente de deslizamiento súbito se requiere la presencia de un gas que esté almacenado a gran profundidad, bien podría ser que este gas sea el detectado en los precursores. En particular, el fenómeno de la formación de llamaradas induce a pensar que se está liberando un gas combustible procedente de una zona profunda, y el candidato más verosímil es el metano no biógeno. No parece probable que a la profundidad a que se encuentran los focos de la mayoría de los terremotos haya gas de origen biológico. Por otra parte, si lo hubiera habido, pronto se habría agotado en las regiones propensas a sufrir movimientos sísmicos.

A menudo se supone que, a gran profundidad, no puede haber material poroso, pues la elevada presión lo estrujaría y haría ascender sus fluidos. Sin embargo, deberíamos admitir que la litosfera tiene dos tipos de porosidad. Al efectuar un sondeo desde la superficie encontramos poros en las rocas, por lo general interconectados y llenos de agua. En los poros la presión es hidrostática (regida por la carga de agua suprayacente) y unas tres veces menor que la presión que soporta la roca adyacente. A medida que se alcanzan niveles más profundos, la porosidad va disminuyendo lentamente; en un momento dado puede alcanzarse un nivel en el cual la roca no tenga ya la resistencia suficiente para soportar la presión diferencial. A este nivel, todos los espacios intersticiales interconectados realmente se cierran, expeliendo hacia arriba cualquier líquido o gas que hubieran podido contener.

Por debajo de este nivel puede haber otro dominio en el cual vuelva a ser factible la existencia de bolsas de gas, pero cuyos fluidos intersticiales estarían sometidos a una presión próxima a la litostática (la presión que experimenta la roca). La profundidad a la cual se pasa de un régimen a otro depende de la resistencia de la roca, pero casi nunca supera los seis kilómetros y puede reducirse a sólo cuatro. Como ambos dominios no pueden tener ninguna conexión permanente, debe haber pues una capa sin poros, a la que llamaremos nivel crítico.

Suponemos que a grandes profundidades por debajo del nivel crítico se genera gas (el volumen del cual varía de una región a otra) a un ritmo lento y constante, bien sea a través de reacciones químicas o bien por difusión hacia afuera de la roca. El gas empieza a abrir poros. Cuando un "dominio poroso" queda interconectado a lo largo de una distancia vertical suficiente (probablemente varios kilómetros), resulta inestable en términos de presión y

empieza a ascender. La razón de ello estriba en que la roca ya no tiene una resistencia suficiente para soportar la diferencia de gradientes de presión entre el gas y la roca más densa.

En esta etapa, los poros del fondo del dominio se cerrarán y se abrirán los de su parte superior; el dominio poroso emigrará lentamente hacia arriba como un todo. Abrirse paso desde una profundidad de varios centenares de kilómetros hasta la superficie probablemente requerirá varios años. Habida cuenta de que, durante el ascenso, disminuye la presión media en el gas, el volumen del dominio poroso deberá aumentar. En este fenómeno puede hallarse la explicación de la elevación gradual de la superficie (unos 10 centímetros) que, en ocasiones, se ha detectado unos cuantos años antes de que sobreviniera un terremoto. Cuando el dominio poroso atraviesa la capa crítica y penetra en la región hidrostática, la rápida apertura de vías de descarga puede provocar efectos precursores y terremotos de foco somero.

Si los precursores macroscópicos y también los que implica dilatancia constituyen verdaderos síntomas secundarios de una causa principal subyacente (la creciente presión del gas y la porosidad en profundidad), cabe la posibilidad de controlar los fenómenos primarios y, con ello, obtener una base más segura de predicción de seísmos. Quizá, mediante una cartografía sísmica de la subsuperficie, sea posible detectar leves cambios en la velocidad de las ondas debidas a las modificaciones de la porosidad; la prospección reiterada y precisa de una misma zona puede realmente permitirnos observar el ascenso de un dominio poroso.

Sería muy útil tratar de controlar directamente la composición y la presión del gas en zonas de falla. En la Unión Soviética existen ya diversos grupos que llevan a cabo programas destinados a observar cambios en la composición química, y en la abundancia relativa de los isótopos de un mismo elemento, de los gases del suelo. Es sabido que ambos tipos de cambios aparecen antes de que ocurra un terremoto. Gran parte de estos estudios los está realizando la cualificada minoría de geoquímicos de la Unión Soviética que también se ha mostrado interesada en la idea del petróleo no biógeno.

Creemos que la hipótesis del gas procedente de las profundidades terrestres, además de sugerir nuevos métodos para la predicción de seísmos, insinúa la posibilidad de que se hayan acumulado vastos volúmenes de metano procedente de fuentes internas en regiones cuya riqueza nunca habríamos sospechado en base a la teoría convencional del origen biológico. El dominio superior, en el cual el gas se encuentra a presión hidrostática, ya se ha explorado bastante profusamente (si bien, quizá, no de forma suficientemente generalizada). Pero el dominio inferior, en el que el gas puede hallarse sólo a presiones más próximas a la presión litostática, apenas lo conocemos. En algunos lugares se ha logrado hacer manar gas a estas presiones, pero siempre se ha interpretado que su presencia allí se debía exclusivamente a una configuración geológica peculiar. Ahora bien, si se tratara de un fenómeno generalizado y debajo del nivel crítico de porosidad cero hubiera otro régimen de gran porosidad, debida a la gran presión del gas, habría que volver a calcular las previsiones de recursos mundiales de combustible. Las cantidades de gas que han estado implicadas en la degasificación carbónica de la tierra han sido enormes. Si el metano ha sido un contribuyente significativo, la fracción del mismo que reside en el dominio de alta presión en su ruta de salida puede ser aún muy grande, en comparación con todos los demás recursos combustibles conocidos.

Los procesos de degasificación de la tierra pueden aclararse mediante muchas otras líneas de investigación. Se pueden observar las variaciones en el contenido en metano de la atmósfera. Es posible controlar los cambios de presión de los fluidos en el suelo. Las leves diferencias en la proporción del isótopo pesado del carbono (carbono 13) entre hidrocarburos de orígenes diferentes revelan una historia complicada que requiere estudios mucho más profundos. Nadie posee pruebas firmes sobre los diversos regímenes gaseosos existentes a profundidades superiores a unos cuantos kilómetros por debajo de la superficie, ni sobre el volumen de los diversos gases que emergen, ni sobre la frecuencia con la que lo hacen.

Sin duda alguna, este intento nuestro de formular una hipótesis relativamente sencilla para explicar muchos hechos que antes no se ponían en relación resultará excesivamente simplista y exagerado en según qué contexto. Sin embargo, albergamos la esperanza de que incite nuevas investigaciones en este campo fundamental de la geofísica y la geoquímica que quizás ayuden a descubrir nuevas fuentes de combustible y, en cualquier caso, a comprender mejor la tierra y sus recursos.

Selección sexual en *Hylobittacus apicalis*

Esta especie ha desarrollado una forma sorprendente de comportamiento de cópula: las hembras eligen a los machos a tenor de la presa que les presentan durante el cortejo, y éstos las imitan para robar dicha presa a los rivales

Randy Thornhill

Los Insectos, a pesar de la relativa simplicidad de su sistema nervioso, muestran diversidad de refinadas y complejas formas de comportamiento de cópula. Pocas actividades de apareamiento son más sorprendentes que las de *Hylobittacus apicalis,* la llamada mosca colgante de puntas negras, un insecto relativamente primitivo con el cuerpo delgado y pardusco, patas largas y finas y cuatro alas estrechas con las puntas negras. Observé por primera vez el comportamiento de esta especie hace nueve años, en un paseo casual a través de una zona boscosa en el sudeste de Michigan. A principios de verano, cuando las poblaciones de moscas colgantes adultas cuentan millares de individuos, no es difícil observarlas volando a través de la vegetación baja o mientras cuelgan de una planta por sus delgadas patas anteriores. Los machos de la especie utilizan sus fuertes patas posteriores para capturar áfidos, moscas domésticas y opiliones falángidos, y pude observar que, una vez un macho había obtenido una de tales presas, comía brevemente y, colgando de una hoja o una ramita, esperaba a que una hembra se le uniera.

Cuando una hembra se posaba junto a un macho, éste le ofrecía su presa y con frecuencia ambos se apareaban. En algunos casos, empero, la hembra reemprendía el vuelo sin más. Las hembras parecían elegir entre los machos, rechazando como pareja a los que las obsequiaban con presas pequeñas y aceptando a los que se las ofrecían grandes. Se me ocurrió que estos insectos podrían proporcionar un excelente sistema en el que probar las ideas de Charles Darwin acerca de la selección sexual, en particular su punto de vista, muy discutido, del papel que en la evolución de la morfología y del comportamiento del macho desempeña la elección que hace la hembra.

Darwin desarrolló su teoría de la selección sexual en *The Descent of Man and Selection in Relation to Sex (El origen del hombre y selección según el sexo),* que se publicó en 1871. Argumentaba que la evolución de caracteres sexuales secundarios podía atribuirse a dos fuerzas selectivas distintas: competencia entre machos por la pareja y elección por parte de la hembra de algunos machos frente a otros. En el primer tipo de selección, la eficacia reproductora de un macho depende de su capacidad de vencer en las luchas con los machos de la misma especie; en el segundo, su eficacia depende de la capacidad de inducir a las hembras exigentes a que lo elijan.

Según Darwin, la evolución de caracteres masculinos tales como las complejas pautas de comportamiento de cortejo, el complicado plumaje de algunas aves, los cuernos de ciertos mamíferos ungulados y las excrecencias corniformes de determinados insectos [véase *Escarabajos cornudos,* por W. G. Eberhard; INVESTIGACION Y CIENCIA, mayo de 1980], todos los cuales comprometen la supervivencia del macho porque le cuestan energía y aumentan su exposición a los depredadores, pueden explicarse en términos de las ventajas que confieren en uno de estos dos procesos selectivos. La importancia de la competencia entre machos considerada como factor de evolución de los caracteres masculinos nunca se ha visto puesta en cuestión seriamente, lo que sí ha ocurrido, y mucho, con la importancia de la elección de la hembra en la selección sexual.

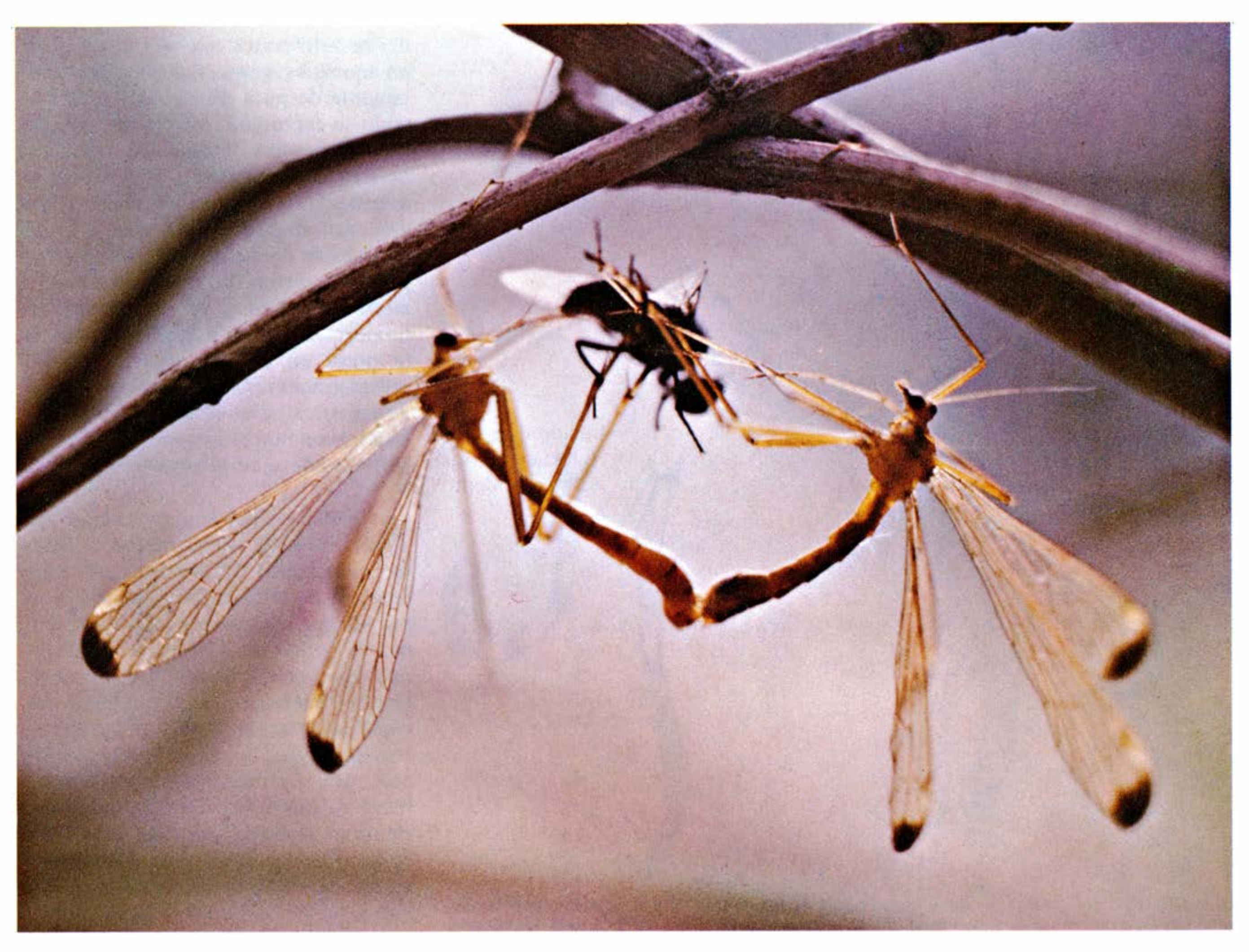
Alfred Russel Wallace, que llegó independientemente de Darwin al concepto de evolución mediante selección natural, fue uno de los primeros en señalar que, para apoyar la formulación de Darwin, simplemente no se disponía de suficientes observaciones de que en la naturaleza tuviera lugar una elección por parte de las hembras. Esta falta de pruebas directas persiste hoy en día. Con mucha frecuencia la existencia de elección femenina se demuestra indirectamente, mediante estudios que indican la distribución de hembras alrededor de machos de edades o tamaños distintos o con cantidades diferentes de un determinado recurso, como el territorio. O bien se supone meramente que la elección de las hembras opera en situaciones en las que éstas se aparean con algunos machos y no con otros, a pesar de que las hembras, que por lo general son más pequeñas, se hallan peor armadas y son menos agresivas que los machos, pueden a veces ser forzadas a permanecer con determinados machos sin tener en cuenta sus propias preferencias. Además, hay pocas pruebas que indiquen que la elección por parte de las hembras sea realmente adaptativa en tales situaciones y sirva para aumentar la supervivencia de aquellas que la practican o, de manera más específica, el número de descendientes que producen.

Desde mi paseo por los bosques de Michigan he venido estudiando el comportamiento de las moscas colgantes y las moscas escorpión (que conjuntamente constituyen el orden de insectos Mecópteros) en los Estados Unidos y en México. Mis hallazgos muestran que las hembras de la especie *Hylobittacus apicalis,* de la familia Bitácidos, valoran efectivamente a los machos en base al obsequio que éstos les presentan

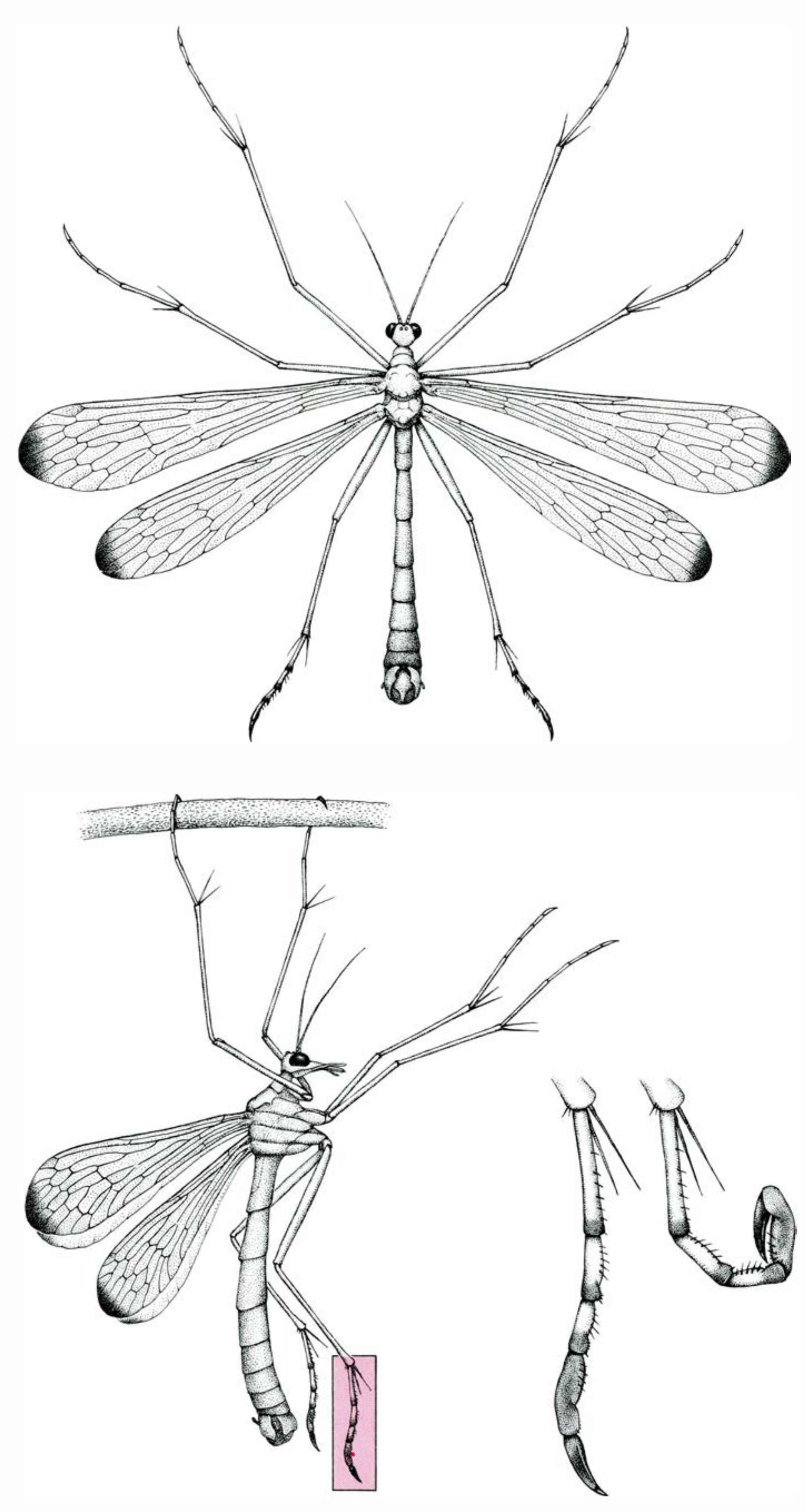
durante el cortejo: una presa que suele ser un insecto. No todas las hembras de este bitácido escogen a los machos así, pero las que lo hacen eligen como pareja a los machos portadores de la presa mayor. Es más, una tal elección aumenta claramente las probabilidades de supervivencia de la hembra y el número de huevos que pone. No hay duda de que el comportamiento de *Hylobittacus apicalis* resulta ser uno de los pocos ejemplos conocidos de elección adaptativa por parte de la hembra. Además, he descubierto que para aumentar el éxito del apareamiento los machos de esta especie compiten por las hembras siguiendo pautas muy curiosas entre las que se incluye la imitación del comportamiento femenino. Por todo ello, la mosca colgante de puntas negras sirve como modelo ideal para comprender la selección sexual tal como Darwin la describió.

Hylobittacus apicalis es asimismo un excelente sujeto para el estudio del comportamiento en condiciones naturales. De unos dos centímetros de longitud, es de actividad diurna y captura a sus presas y se aparea durante todas las horas del día. Puede marcarse a los individuos aplicando una mancha de pintura a su cuerpo o alas, y aunque machos y hembras se parecen, pueden distinguirse por sus pautas de vuelo: los machos vuelan rápidamente de un lado para otro, mientras que las hembras, sobrecargadas por el peso de los huevos, vuelan de manera regular y en línea recta. Ambos sexos son relativamente lentos en su desplazamiento, realizando sólo vuelos cortos a través del estrato herbáceo de la vegetación de su ambiente forestal, de manera que pueden seguirse fácilmente en su frecuente interacción con otras moscas colgantes y con los insectos y otros artrópodos de los que se alimentan.

La secuencia de apareamiento de *Hylobittacus apicalis* se inicia cuando un macho captura un artrópodo, o se lo roba a otro macho o a una pareja apareada y, manteniendo a la presa en sus tarsos (los últimos artejos de las patas) posteriores prensiles, se agarra a una hoja o a una ramita y comienza a comer. (Para matar a la presa el bitácido utiliza su penetrante probóscide; cuando esta estructura en forma de pico penetra en el cuerpo de la presa, se liberan enzimas que la paralizan y licúan sus contenidos corporales, de manera que puedan ser succionados.) No todas las presas que los machos capturan se ofrecen a las hembras como regalo nupcial. De hecho, mis observaciones de 42 machos marcados en dos localidades de estudio en los bosques del sudeste de Michigan revelan que los machos sopesan mucho qué presa emplearán a este fin. De los 345 insectos que se observó que los machos marcados capturaban, 110, es decir,



APAREAMIENTO DE UNA PAREJA de moscas colgantes de puntas negras (*Hylobittacus apicalis*); ambos insectos agarran el regalo nupcial (una gran moscarda) que el macho (*izquierda*) ofrece a la hembra (*derecha*). Estos bitácidos se hallan en la primera fase del proceso de apareamiento, en el que la hembra valora la presa que le ofrece el macho y éste intenta aparearse con aquella. La hembra de *Hylobittacus apicalis* juzga a los machos según el tamaño de la presa que éstos le ofrecen, y rehúsa aparearse, o lo hace sólo brevemente, con los machos que le agasajan con presas pequeñas o incomestibles. Si la presa que se ofrece es suficientemente grande (como ocurre aquí), la hembra acepta al macho y copula con él 20 minutos o más.



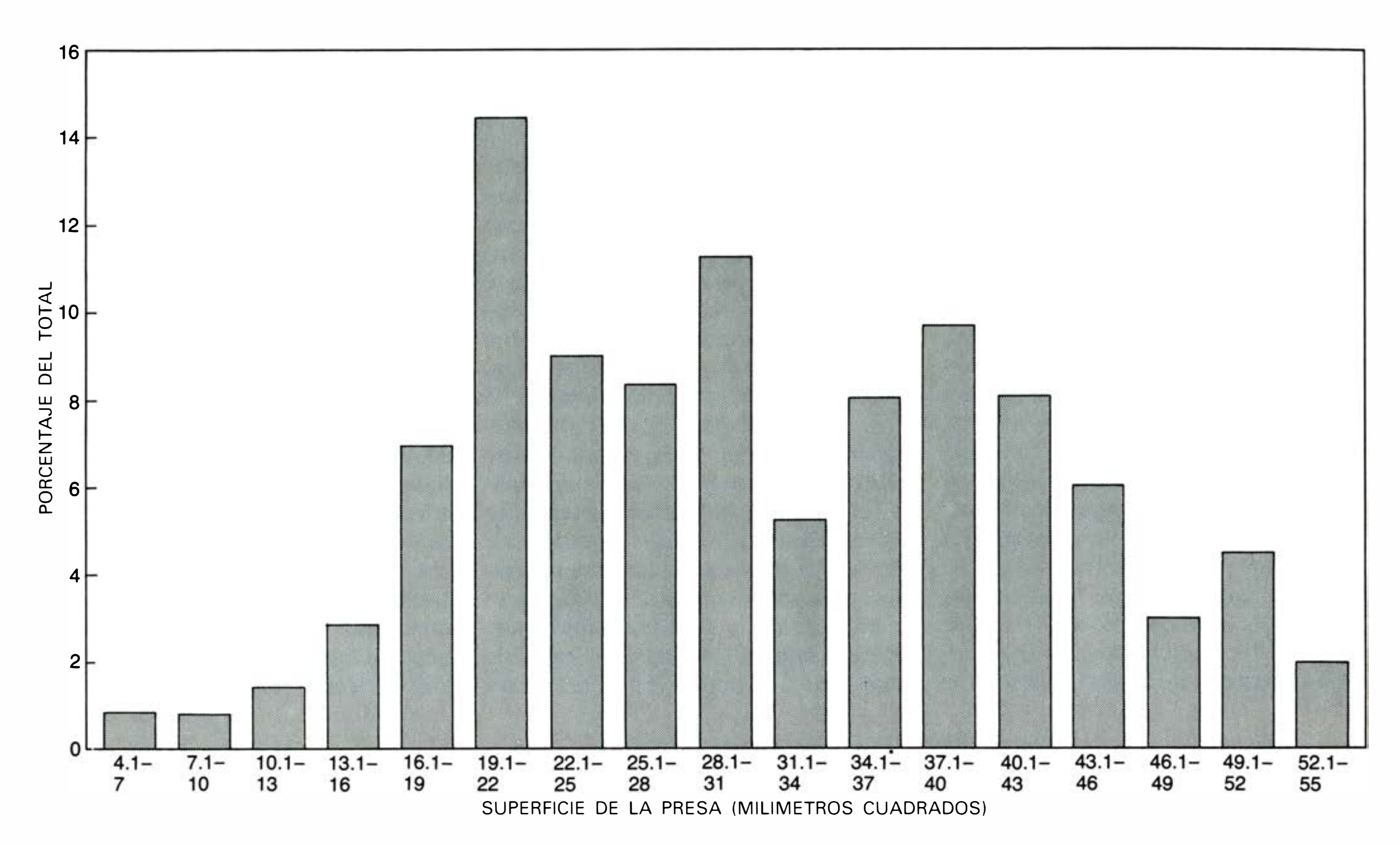
HYLOBITTACUS APICALIS, que aquí se ilustra en sus aspectos dorsal y lateral, es un insecto de tamaño medio (de unos dos centímetros de longitud) que se encuentra entre la vegetación baja de las zonas boscosas de los Estados Unidos orientales. Cuando el insecto no vuela, pende de sus largas y delgadas patas anteriores de una hoja o de una ramita, como se indica abajo, a la izquierda. Para capturar y manipular a su presa la mosca colgante de puntas negras (según se la conoce por dicho hábito y por el color de sus alas) utiliza sus robustos y prensiles tarsos (*color*), los últimos segmentos articulados de sus patas posteriores. Como se indica en la ilustración aumentada de abajo, a la derecha, los últimos cuatro artejos de cada tarso proporcionan un poderoso órgano para envolver a la presa; los dos últimos artejos actúan de tenazas para agarrar una pata o un ala. *H. apicalis* es un insecto relativamente primitivo.

cerca del 32 por cien, fueron abandonados en la hojarasca del suelo del bosque después de que los machos se alimentaran brevemente de ellos. Todas las presas desechadas medían menos de 16 milímetros cuadrados de superficie, lo que sugiere que eran demasiado pequeñas para constituir ofrendas adecuadas. (En las investigaciones de las que estoy hablando aquí se considera que la superficie de un artrópodo es el producto de su anchura por su longitud; por ejemplo, una mosca doméstica de tamaño medio tiene una superficie de unos 20 milímetros cuadrados.)

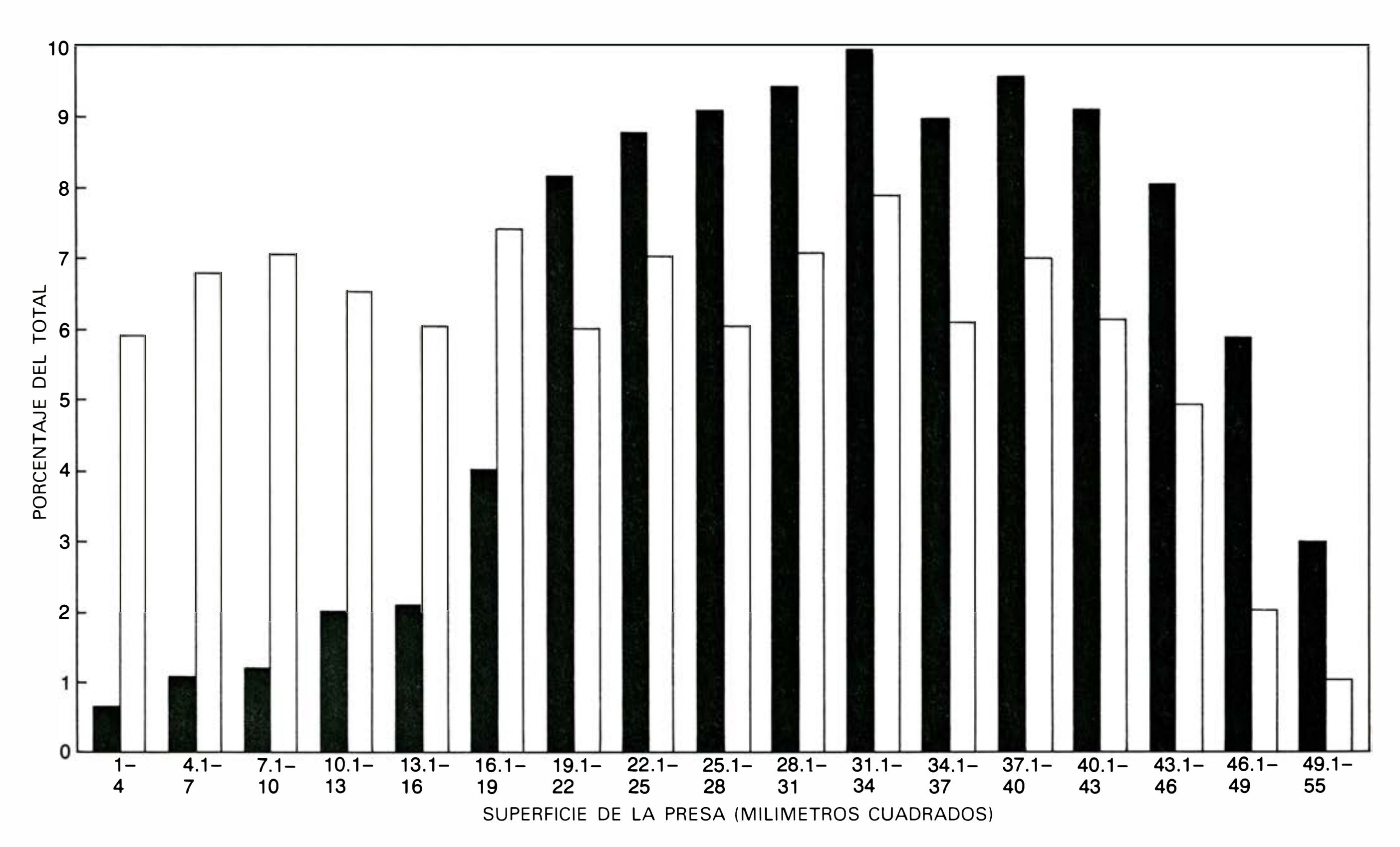
Se obtuvo una prueba adicional de que los machos retenían las presas en base a su tamaño del siguiente modo: la vegetación herbácea de cada una de las localidades de mi estudio se barría con una red para insectos, o cazamariposas. Se tomaban muestras de tal barrido cada día durante una semana para determinar la distribución de tamaños de los artrópodos que se encontraban en aquellos lugares. Además, inmediatamente después de efectuar cada barrido se capturaban con la red parejas de *Hylobittacus apicalis* en cópula y machos con una presa, y se medía el tamaño de éstas. La comparación entre estas muestras demostró que, aunque los machos de nuestro bitácido sujetan presas pequeñas (de menos de 16 milímetros cuadrados de superficie) en proporción directa a la abundancia de la presa en la vegetación, capturan presas grandes (de 16 milímetros o más de superficie) y las ofrecen como regalo nupcial con mayor frecuencia de lo que corresponde a su abundancia natural.

A mediados del verano, cuando los machos de *Hylobittacus apicalis* son abundantes, raramente se ve a las hembras de la especie capturar presas por su cuenta; dependen casi exclusivamente de los machos y de las presas que éstos proporcionan. Las observaciones demuestran que cuando las hembras cazan no desechan las presas más pequeñas, y mis estudios confirman que las hembras, consideradas fuera del contexto del apareamiento, capturan distintos tamaños de presas en proporción directa a la distribución natural de tamaños de las mismas. Así pues, son los machos los depredadores selectivos, y obtienen con efectividad presas de determinados tamaños dentro de la gama disponible.

Cuando una mosca colgante macho ha capturado una presa de tamaño suficiente, prueba unos bocados y,



LOS MACHOS PREFIEREN PRESAS GRANDES para ofrecerlas como obsequio nupcial. El histograma indica la distribución de tamaños de los artrópodos empleados como presa por 168 machos de *Hylobittacus apicalis* en su apareamiento. Las observaciones de esta especie en su ambiente natural señalan que más del 90 por ciento de todas las cópulas implican presas grandes, es decir, insectos u otros artrópodos que midan 16 milímetros cuadrados o más de superficie. (Para el propósito de estas investigaciones, la superficie de un artrópodo se calcula como el producto de su longitud por su anchura; por ejemplo, la moscarda que sostienen los bitácidos de la ilustración de la página 99 mide unos 30 milímetros cuadrados de superficie.

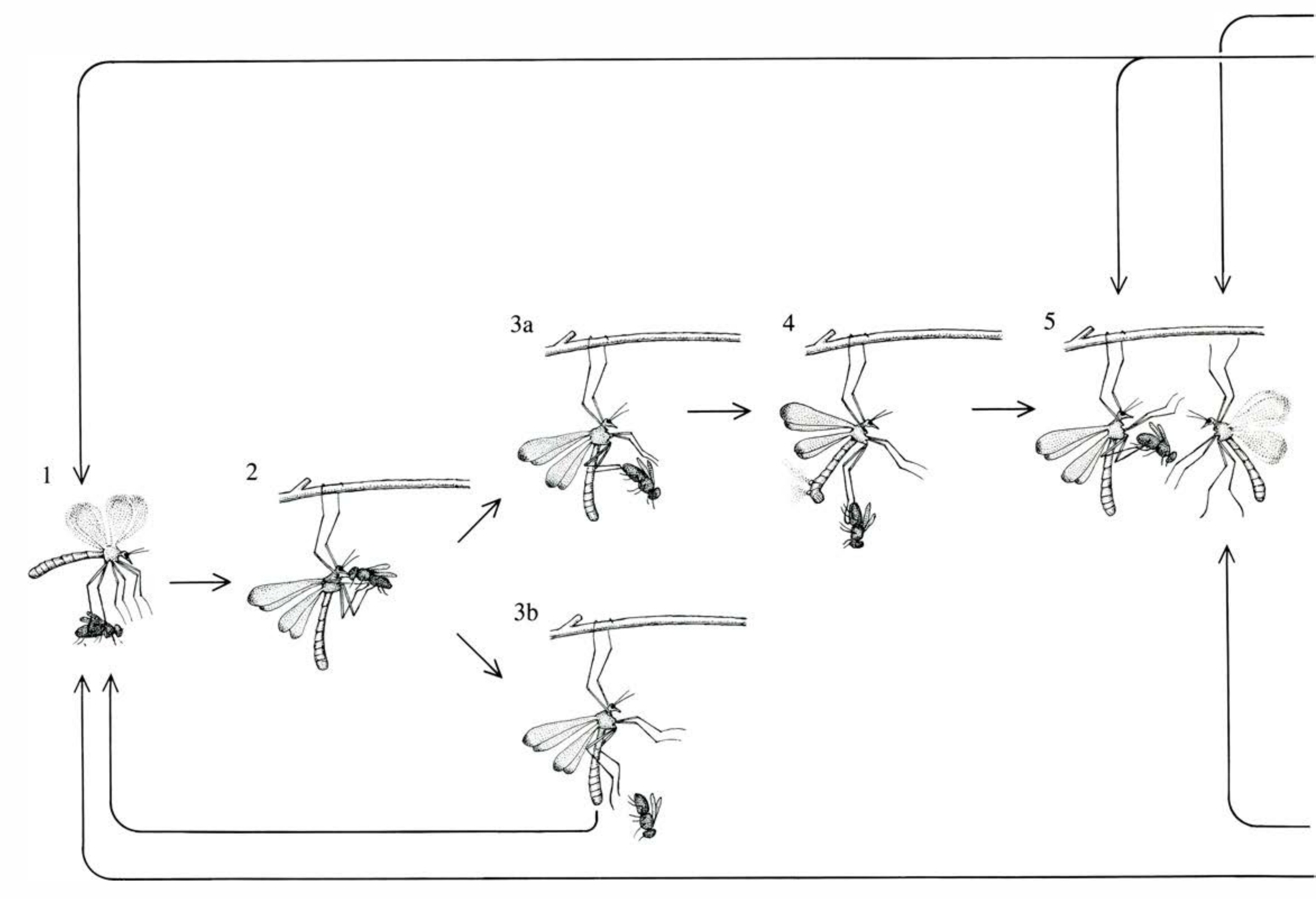


CUANDO LAS HEMBRAS CAZAN no optan por presas grandes, como se desprende de este histograma que compara la distribución de tamaños de las presas obtenidas de 248 machos solos (*negro*) y de 276 hembras solas (*blanco*) de *Hylobittacus apicalis* en las localidades forestales de estudio. Las hembras de este bitácido, suelen estar bien aprovisionadas de alimento por los machos que buscan pareja, por lo que generalmente no cazan.

sujetándola con sus fuertes tarsos posteriores, realiza cortos vuelos a través de la vegetación herbácea. En los machos marcados que observé, estos vuelos prenupciales raramente excedían de los 3,6 metros. Las señales que inicialmente atraen a una hembra a las cercanías del macho son olfativas, no visuales. Al final de cada vuelo prenupcial el macho se cuelga de una hoja o una ramita, evagina un par de sacos glandulares de su abdomen y emite una feromona, es decir, una sustancia portadora de información que sirve para atraer a las hembras. (Las glándulas de la feromona son evaginadas mediante un aumento de la presión sanguínea en el abdomen del macho, y se retraen por la acción de músculos especializados.) El apareamiento no empieza realmente hasta que una hembra se halla colgando de sus patas anteriores frente al macho. El macho presenta entonces su regalo nupcial a la hembra y, mientras sigue sujetándolo, permite que ésta lo agarre con sus tarsos posteriores. Cuando la hembra comienza a alimentarse el macho intenta copular con ella.

Si el regalo del macho tiene el tamaño adecuado (16 milímetros cuadrados o más de superficie), la hembra acepta al macho y los dos copulan durante una media de 23 minutos, tiempo durante el cual la hembra sigue comiendo. El proceso de apareamiento lo finaliza el macho; se desencadena entonces una lucha mientras éste intenta zafarse de la hembra y arrancar la presa que ella sujeta. Pero si el regalo del macho no es de su agrado (como ocurre cuando se trata de una mariquita, es decir, un coleóptero coccinélido que contiene sustancias repugnantes) o demasiado pequeño, la hembra rehusará copular con el macho. Si la cópula ya se ha iniciado, la finalizará rápidamente, marchándose volando al cabo de unos cinco minutos. En este último caso, pues, es la hembra la que intenta zafarse del macho. Ahora bien, si éste gana en la lucha subsiguiente por la presa, tratará por lo general de presentar su regalo "indeseable" a la hembra y atraerla de nuevo a la cópula.

Se observó que los machos vencían en el 64 por ciento de las luchas entre parejas de moscas colgantes por la posesión final del regalo nupcial, y las hembras sólo en el 8 por ciento de los envites. En los restantes casos la presa cayó y se perdió en la hojarasca del suelo del bosque. Si el macho consigue retener la presa una vez terminada la cópula, come brevemente de ella y, si todavía puede proporcionar alimento adecuado, la utiliza de nuevo para repetir la secuencia de apareamiento con una hembra distinta. En caso contrario, la desecha e inicia la caza de otra presa apta para la alimentación nupcial. De su tamaño depende que la presa pueda o no ser empleada en dos



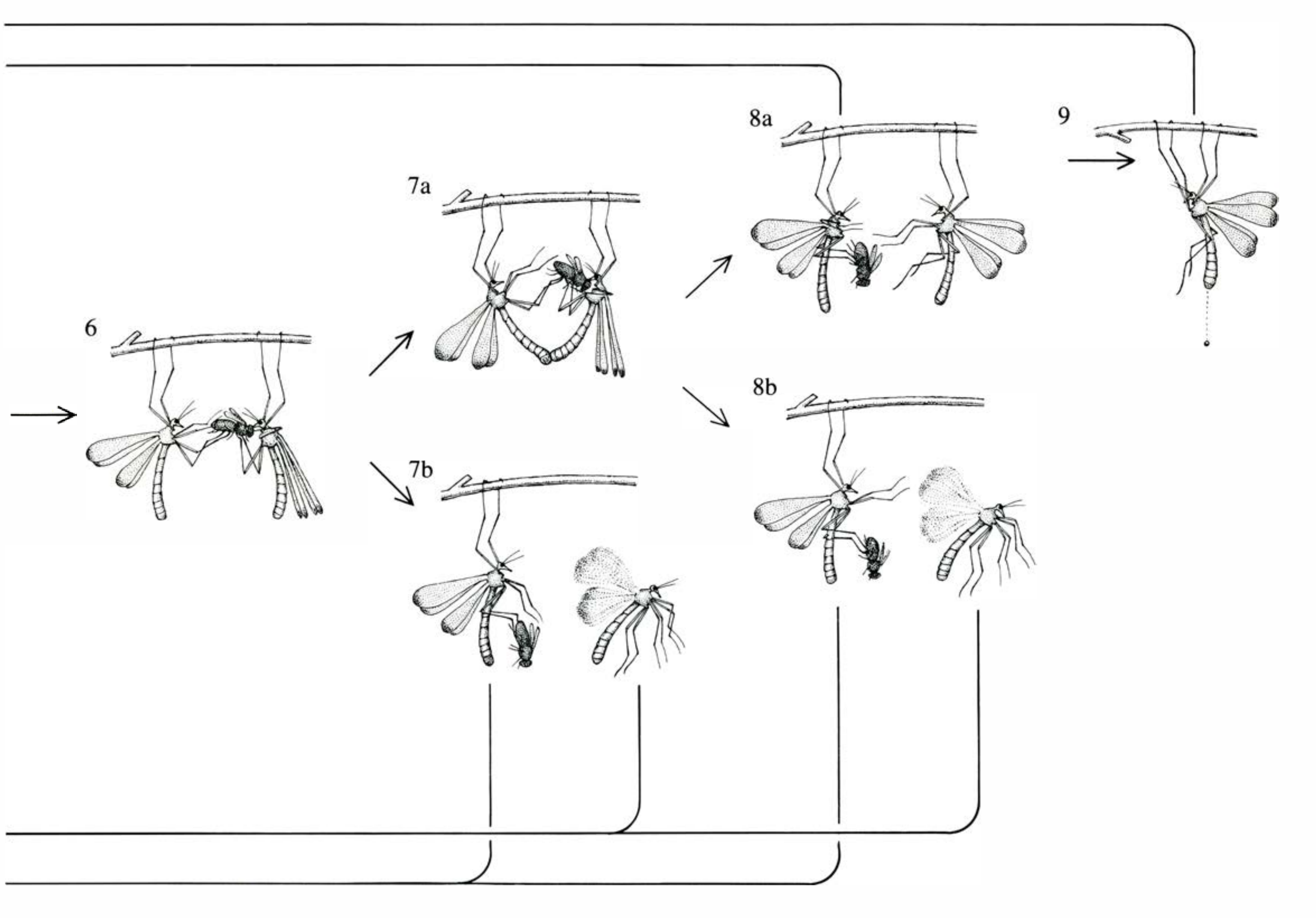
SECUENCIA DE APAREAMIENTO, que comienza cuando un macho de mosca colgante de puntas negras captura un artrópodo o bien roba una presa de una pareja apareada o de otro macho (*1*); después, mientras pende de una hoja o una ramita, se alimenta brevemente de ella (*2*). Si la presa es demasiado pequeña, el macho suele desecharla en la hojarasca del suelo del bosque (*3b*), en cuyo caso debe obtener una nueva presa y empezar de nuevo. Sin embargo, si la presa es suficientemente grande, el macho la retiene (*3a*) y, evaginando un par de sacos glandulares de su abdomen, emite una feromona, es decir, una sustancia portadora de información que sirve para atraer a las hembras (*4*). Cuando llega una hembra de *Hylobittacus apicalis*, el macho le ofrece la presa (*5*), permitiéndole que la agarre con sus tarsos posteriores mientras él la sigue manteniendo sujeta (*6*). La hembra come de la presa, sopesándola mientras el macho se dispone a copular. Si la presa no es aceptable (porque es incomestible o demasiado pequeña) la hembra puede soltarla y

apareamientos sucesivos. He comprobado que los artrópodos que miden entre 28 y 45 milímetros cuadrados de superficie son las presas que con más frecuencia se utilizan dos veces. Las presas menores no proporcionan suficiente alimento, y las de tamaño superior al indicado son difíciles de sujetar por el macho en la lucha postcopulatoria. De hecho, son las hembras quienes tienen más probabilidad de retener una presa ofrecida como regalo nupcial si ésta es de gran tamaño.

Así pues, la hembra de *Hylobittacus apicalis* puede mostrar su resistencia a los machos con presas pequeñas a dos niveles: en primer lugar, rechazando a tales machos antes de empezar la cópula y, segundo, aceptándolos como pareja sólo durante un período limitado. Los machos con presas pequeñas no consiguen aparearse por lo general por más de cinco minutos, mientras que los que ofrecen presas grandes copulan durante 20 minutos o más. El significado de estas cifras no quedó claro hasta que realicé experimentos de laboratorio para determinar en qué momento, en el curso de la cópula, se transfiere realmente el esperma del macho a la hembra. En estos experimentos se hacía copular a hembras vírgenes durante períodos que iban de uno a 39 minutos. A continuación se extraía su órgano almacenador de esperma y se aplicaba una tinción específica para los espermatozoides a fin de poderlos contar.

Estos estudios demostraron que eran necesarios cinco minutos de cópula para que se transfiriera algún espermatozoide al órgano de almacenamiento de la hembra. Además, en los apareamientos que duraron entre cinco y 20 minutos hubo una correlación positiva entre la duración de la cópula y el número de espermatozoides transferidos. Pasados 20 minutos había poco aumento en el número de espermatozoides transferidos, independientemente de por cuánto tiempo continuara todavía la cópula. De ahí que los breves apareamientos de los machos con presas de tamaño inferior al requerido dieran lugar a transferencia de espermatozoides una reducida, o bien ninguna, y por ello a un pequeño aumento del número de descendientes engendrados por estos machos, o a ninguno. En cambio, las cópulas prolongadas de los machos que sostenían presas grandes provocaban una inseminación máxima de la hembra, con las obvias consecuencias en la producción de descendientes.

La máxima inseminación de la hembra no es la única ventaja que el macho con la presa mayor obtiene de su prolongado período de cópula. Observaciones de *Hylobittacus apicalis* tanto en el laboratorio como en el campo



marcharse volando, rehusando el apareamiento solicitado (*7b*). O bien la hembra puede copular con el macho, pero únicamente por un breve período, terminando la interacción al cabo de unos cinco minutos, por término medio (*8b*). En cambio, si la presa es suficientemente grande, la pareja de moscas colgantes copulará (*7a*) durante al menos 20 minutos por término medio, en cuyo caso es el macho el que da por terminada la interacción cuando le parece (*8a*). La hembra come durante todo el período de cópula, que invariablemente finaliza en una lucha por la posesión de la presa nupcial. (Por lo general, el macho gana este combate, y, si la presa es suficientemente grande, la emplea por segunda vez como ofrenda nupcial a otra hembra.) Una hembra de *Hylobittacus apicalis* que ha copulado sólo durante cinco minutos empieza de nuevo el proceso de apareamiento. Una hembra que ha copulado 20 minutos o más, se muestra indiferente a cualquier insinuación de apareamiento por unas cuatro horas, durante las cuales pone una media de tres huevos (9).

muestran que el período de cópula más largo estimula la oviposición y hace asimismo menor la probabilidad de que la hembra vuelva a aparearse pronto. Las moscas colgantes de puntas negras hembras ponen huevos grandes y visibles, que se adhieren a ellas durante unos cuantos minutos antes de caer al suelo del bosque. Gracias a este tipo de oviposición fue posible determinar la producción de huevos de las hembras después de su interacción con machos que ofrecían presas de tamaños distintos.

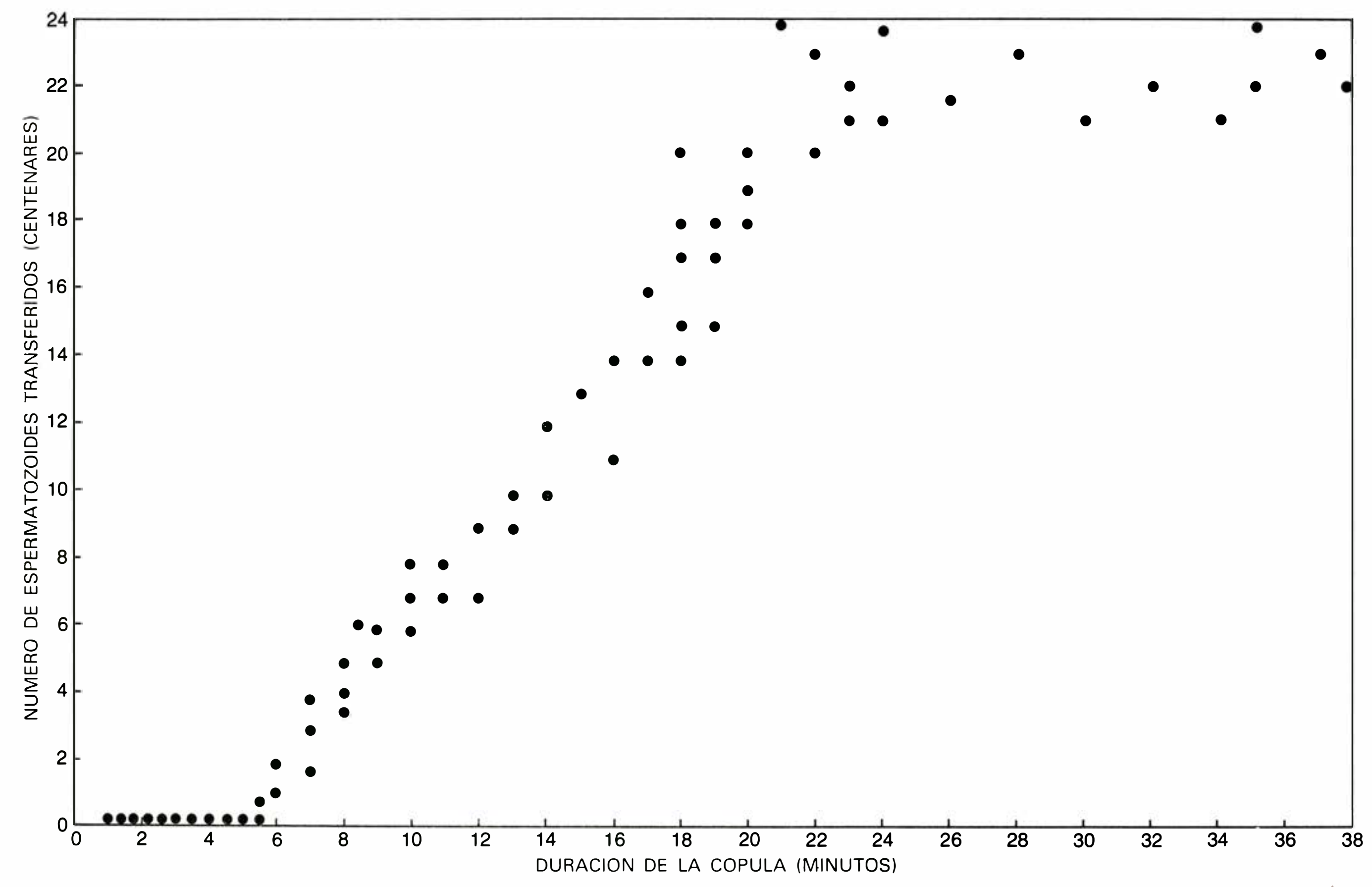
He comprobado que, tras copular durante 20 minutos o más, la hembra se muestra indiferente a los machos y no vuelve a aparearse hasta pasadas unas cuatro horas, tiempo durante el cual pone una media de tres huevos. Si continúa encontrando machos con regalos nupciales grandes, este ciclo de no receptividad para el apareamiento y de oviposición puede continuar sin interrupción durante todo el día. Por otra parte, si una hembra ha copulado únicamente unos cinco minutos, no pone ningún huevo, sino que toma contacto rápidamente con otro macho y después otro, buscando siempre individuos portadores de una presa grande.

No importa lo rico que sea el suministro de nutrientes que la hembra obtenga de las presas pequeñas en el curso de estas cortas interacciones: sigue interactuando brevemente con distintos machos hasta que encuentra uno con una presa adecuada; entonces emprende una cópula prolongada.

La elección que efectúa la hembra de *Hylobittacus apicalis* es claramente adaptativa, es decir, aumenta la capacidad reproductora de aquellas hembras que la practican. Descubrí que las hembras que discriminaban entre machos en base a los ofrecimientos nupciales ponían significativamente más huevos por unidad de tiempo que las que no discriminaban. Además, las hembras que discriminaban aumentaban su propia probabilidad de supervivencia.

La razón es la siguiente. Un bitácido que vuele a través de su ambiente forestal en busca de caza corre el riesgo de acabar atrapado en una telaraña. De hecho, una fracción sustancial de la población de moscas colgantes se pierde debido a la depredación que sobre ellas efectúan las arañas tejedoras de telas. Sin embargo, entre las moscas colgantes de puntas negras la tasa de pérdidas es significativamente más baja para las hembras que para los machos, fenómeno que puede atribuirse a la disparidad en las distancias recorridas por los dos sexos en el curso de sus actividades normales. En los individuos marcados observé que los machos se desplazaban por término medio unos 33 metros cada hora, mientras que para las hembras la distancia media recorrida era de unos 16 metros.

El origen de esta diferencia puede buscarse en la práctica de la alimentación nupcial: los machos pasan alrededor del 50 por ciento de su tiempo entre cópulas cazando presas nupciales, mientras que las hembras, que están bien abastecidas de alimento por los machos, raramente precisan cazar. Por ello es mucho menos probable que las hembras encuentren telarañas. Al seleccionar un macho con una presa grande, la hembra limita todavía más su movimiento, y con ello el riesgo de ser capturada por una araña tejedora, de manera que aumenta sus probabili-



NUMERO DE ESPERMATOZOIDES transferidos desde el macho de *Hylobittacus apicalis* al órgano de almacenamiento de esperma de la hembra. Dicha cifra depende de la duración de la cópula. Como se indica en este gráfico que muestra los resultados de 66 apareamientos de distinta duración, en el órgano de almacenamiento de la hembra prácticamente no entra esperma en los primeros cinco minutos de la cópula. Transcurrido ese intervalo, hay una relación directa entre la duración de la cópula y el número de espermatozoides que se transfieren. Sin embargo, para apareamientos que se prolongan más de unos 20 minutos, el número medio transferido permanece alrededor de unos 2200 espermatozoides. Los machos que copulan durante 20 minutos o más son los que ofrecen presas grandes, de esta manera, únicamente estos machos son los que lograrán la inseminación máxima de las hembras.

dades de supervivencia. La relación entre alimentación nupcial y supervivencia femenina en *Hylobittacus apicalis* viene soportada por una comparación con otro bitácido, *Bittacus strigosus*, que no realiza alimentación nupcial. En esta especie, machos y hembras participan en la caza de manera aproximadamente igual, y las tasas de mortalidad debidas a depredación por arañas son comparables en los dos sexos.

Mis hallazgos sugieren que la alimentación nupcial evolucionó en la mosca colgante de puntas negras como un medio de reducir la participación de la hembra en actividades peligrosas, a saber, las asociadas con la caza. Al suministrar alimento a la hembra, el macho reduce la probabilidad de que ésta sea capturada por una araña antes de poner los huevos que él ha fecundado. Debe señalarse que mi interpretación de la ventaja selectiva de la alimentación nupcial para los machos se basa en la suposición, respaldada por estudios de otros insectos, de que el último macho que copula con una hembra durante 20 o más minutos fecunda la mayoría de los huevos que ésta pone.

El hecho de que la elección que hace la hembra de *Hylobittacus apicalis* signifique que algunos machos (los que tienen la presa grande) se apareen con más frecuencia y éxito que otros (los que tienen la presa pequeña) indica que esta elección es en potencia una forma importante de selección sexual, que posiblemente excluye de la reproducción a una fracción considerable de machos en cada generación de este bitácido. Mis estudios indican que, debido fundamentalmente al limitado número de presas disponibles, sólo un pequeño porcentaje, entre el 2 y el 10 por ciento, de los machos adultos de una población de moscas colgantes se hallan en posesión de una presa en un momento dado.

De estos machos, puede considerarse que los que han sido vistos realizando los vuelos prenupciales característicos o liberando feromonas han evaluado a sus presas con fines de apareamiento. En cualquier momento dado, el 90 por ciento de los machos que han evaluado positivamente las presas están en posesión de presas grandes. En otras palabras, el 10 por ciento de los machos que han valorado las presas estarán en posesión de presas pequeñas, y por lo tanto serán discriminados en contra por las hembras porque su presa no será adecuada. De ahí que la elección de las hembras puede excluir de la reproducción hasta un 10 por ciento de todos los machos adultos de la población. Además, la variación en la capacidad de los machos de encontrar rápidamente presas grandes impide que otros muchos machos alcancen su potencial reproductor máximo.

Es importante comprender que el descubrimiento de que la elección que hace la hembra entre las moscas colgantes de puntas negras produce una variación no aleatoria en la reproducción de los machos, indica únicamente el potencial para la evolución por medio de la selección sexual. Para demostrar que la elección de la hembra produce realmente evolución, que se define como cambios en las frecuencias génicas de una población, debe demostrarse que los machos elegidos por las hembras difieren genéticamente de los no elegidos. Parece que éste es el caso: mi trabajo indica que en la naturaleza los machos que seleccionan presas grandes lo siguen haciendo una y otra vez a lo largo del tiempo, y lo mismo puede decirse de los machos que seleccionan presas pequeñas.

Uno de los métodos que los machos de *Hylobittacus apicalis* emplean para reducir su movimiento y, con ello el riesgo de depredación por las arañas, es el de robar presas a otros machos o a parejas en cópula. El robo de la presa es un componente importante de la competencia entre machos para conseguir pareja, que es el tipo más familiar de selección sexual, y la reducción del riesgo no es la única ventaja que este comportamiento confiere. He comprobado que un macho que roba una presa para destinarla a regalo nupcial pasa por término medio un 42 por ciento menos de tiempo en cada secuencia de apareamiento, de manera que su frecuencia de cópula es significativamente superior a la de un macho que sólo captura sus presas. Asimismo, para reproducirse con éxito, un macho al que se le ha robado la presa debe obtener otro obsequio nupcial, y hay una correlación positiva entre el número de veces que un macho es objeto de robo y el tiempo medio que pasa entre apareamientos sucesivos.

Las observaciones de machos marcados entre cópulas también indican que cualquier macho puede cazar o bien robar una presa. El modo por el que un determinado macho obtiene su presa nupcial depende de lo que encuentra primero: un artrópodo vivo o un macho o una pareja en cópula en posesión de una presa. Esta estrategia de caza mixta se desarrolló probablemente porque el éxito de un macho en hurtar una

presa depende de la presencia de otros machos, que a su vez depende del tamaño de la población de bitácidos y de su proporción sexual, dos factores que fluctúan en el tiempo y en el espacio. Cuando un macho intenta robar una presa utiliza uno de estos dos métodos: o bien arremete volando contra el macho o la pareja en cópula que poseen la presa y se enzarza en una lucha para obtener la posesión de la misma, o bien actúa como una hembra para seducir a otro macho con el fin de que éste le entregue su presa.

En multitud de animales, entre los cuales figuran hienas, muflones de Dall, aves, salamandras, peces y primates, se han observado conductas en que un macho parece imitar a una hembra. Es indiscutible que, en *Hylobittacus apicalis,* el macho exhibe un verdadero mimetismo al imitar a la hembra. Al éxito de este mimetismo coadyuvan varios factores. Machos y hembras son aproximadamente del mismo tamaño y color, y los genitales externos del macho no son tan grandes o tan aparentes como los de los machos de otras especies de bitácidos. Además, el mimetismo que de las hembras hacen los machos en la mosca colgante de puntas negras es claramente adaptativo: los machos que lo practican capturan más rápidamente a las presas, de manera que aumenta su frecuencia de cópula, y también se desplazan menos y probablemente están sujetos a una depredación menor por parte de las arañas constructoras de telas.

Los machos tienen la misma probabilidad a lo largo de todo el proceso de apareamiento a que otros machos les roben la presa, pero es precisamente en el momento en que un macho libera su feromona cuando puede ser objeto de robo por un macho que imita a una hembra, pues está "esperando" la llegada de una hembra. Puesto que sólo ésta responde a la feromona, un ladrón de presas en potencia es atraído hacia el macho que libera la feromona no por señales olfativas, sino mediante señales visuales, pero únicamente en el contexto de la liberación de feromona es apropiado el mimetismo de la hembra.

Para imitar el comportamiento de una hembra, el macho vuela hasta un bitácido que está liberando feromona y hace descender sus alas. El intento de robo de una presa es la única actividad (aparte del descanso nocturno) en la que un macho abate las alas. El abatimiento de las alas por parte de la hembra constituye para el macho una señal de que está preparada para iniciar el proceso de apareamiento. Se trata pues de una señal para que el macho ofrezca a la hembra su regalo nupcial. Sin embargo, cuando un macho que imita a una hembra abate sus alas, el macho emisor de feromona responde huyendo y volando, alrededor de la tercera parte de las veces, aparentemente porque el mimetismo no era convincente. Si no es así, el macho presenta su regalo nupcial al mimeta. Cuando ambos machos sujetan la presa, el mimeta empieza a comer mientras el dueño intenta copular con él. El imitador mantiene alejado su abdomen del alcance de los genitales del otro macho, exactamente igual que hace la hembra durante el período inicial de alimentación y evaluación de la presa. Transcurridos unos dos minutos, el dueño intenta arrancar la presa del abrazo del mimeta. El 22 por ciento de las veces el esfuerzo no tiene éxito y el imitador levanta el vuelo con la presa. (Sólo en alrededor de la mitad de los intentos de robar la presa de los machos que emiten feromona se utiliza el mimetismo de la hembra. En los restantes casos, el ladrón agarra directamente la presa, táctica que tiene éxito en el 14 por ciento de las ocasiones.)

En muchas especies animales existe competencia entre los machos por recursos que se precisan para atraer a las hembras: territorio, lugares de anidación, alimento, etcétera. De ahí que los fenómenos que acabo de describir referentes a la elección de machos por parte de la hembra y a la competencia por las hembras que realizan los machos del bitácido *Hylobittacus apicalis* pudieran ser aplicables al comportamiento de apareamiento de gran variedad de animales, tanto invertebrados como vertebrados. Además, las especies actuales de Mecópteros, el pequeño orden de insectos al que pertenecen las moscas colgantes, son relictos de un orden que antaño fue grande y estuvo ampliamente distribuido. Los estudiosos de la evolución de los Insectos creen que una antigua especie de mecóptero dio origen a varios de los órdenes superiores de insectos, en concreto a los Lepidópteros (las mariposas y polillas), los Dípteros (las moscas verdaderas) y los Sifonápteros (las pulgas). Entender las presiones evolutivas que modelaron el comportamiento reproductor de los Mecópteros debiera conducir a una mejor comprensión del comportamiento reproductor en general.

Juegos matemáticos

La captura del monstruo matemático: un grupo algebraico con un número terrible de elementos

Martin Gardner

Los especialistas de teoría de grupos (una de las ramas del álgebra abstracta) de todo el mundo han estado bregando durante los seis últimos años para atrapar un grupo apodado "el monstruo". El mote está justificado por su colosal tamaño: el número de elementos del monstruo es 808.017.424.794.512.875.886.459.904. 961.710.757.005.754.368.000.000.000, o sea, $2^{46} \times 3^{20} \times 5^{9} \times 7^{6} \times 11^{2} \times 13^{3} \times 17 \times 19 \times 23 \times 29 \times 31 \times 41 \times 47 \times 59 \times 71$. De todas formas, Robert L. Griess Jr., matemático de la Universidad de Michigan que consiguió echarle el lazo al grupo en enero de este año, prefiere darle al monstruo las señas de identidad que propiamente le pertenecen: F_1. Griess (pronunciado "gráis"), que actualmente se encuentra en licencia sabática en el Instituto de Estudios Avanzados de Princeton, demostró la existencia del monstruo construyéndolo efectivamente. La noticia ha sido muy estimulante para los cultivadores de la teoría, pues los aproxima al fin de una tarea emprendida hace más de un siglo: clasificar exhaustivamente todos los grupos.

Una detonación fue la señal de partida para tan ambiciosa empresa. En 1832, un pistoletazo ponía fin a los días de Evariste Galois, joven genio matemático francés, de ideas políticas radicales, muerto en estúpido duelo por un asunto de faldas. Aún no había cumplido los 21 años. Aunque ya se habían realizado trabajos preliminares y fragmentarios sobre el concepto de grupo, fue Galois quien puso cimientos a la moderna teoría, e incluso quien le dio nombre, todo ello, en una larga y amarga carta a un amigo, escrita la noche antes de su muerte.

¿A qué llamamos grupo? A grandes rasgos, es un conjunto de operaciones a efectuar sobre algo, con la propiedad de que si a continuación de una cualquiera de ellas se ejecuta otra de las operaciones de la colección, el resultado final puede en todos los casos conseguirse mediante una sola operación del conjunto. Las operaciones se denominan elementos del grupo, y su número es el orden del grupo.

Antes de dar una definición más rigurosa, pondremos un ejemplo. Se encuentra Vd. en posición de "firmes", dispuesto a obedecer una cualquiera de estas órdenes: "firmes", "izquierda", "media vuelta" y "derecha". Supongamos ahora que ejecute Vd. "izquierda", y a continuación, "media vuelta". Diremos que dos operaciones se "multiplican" al ejecutarlas sucesivamente. Notemos que el "producto" de esta multiplicación concreta puede obtenerse mediante una sola operación: "derecha". La colección de estas cuatro operaciones es un grupo, porque satisface los cuatro axiomas siguientes:

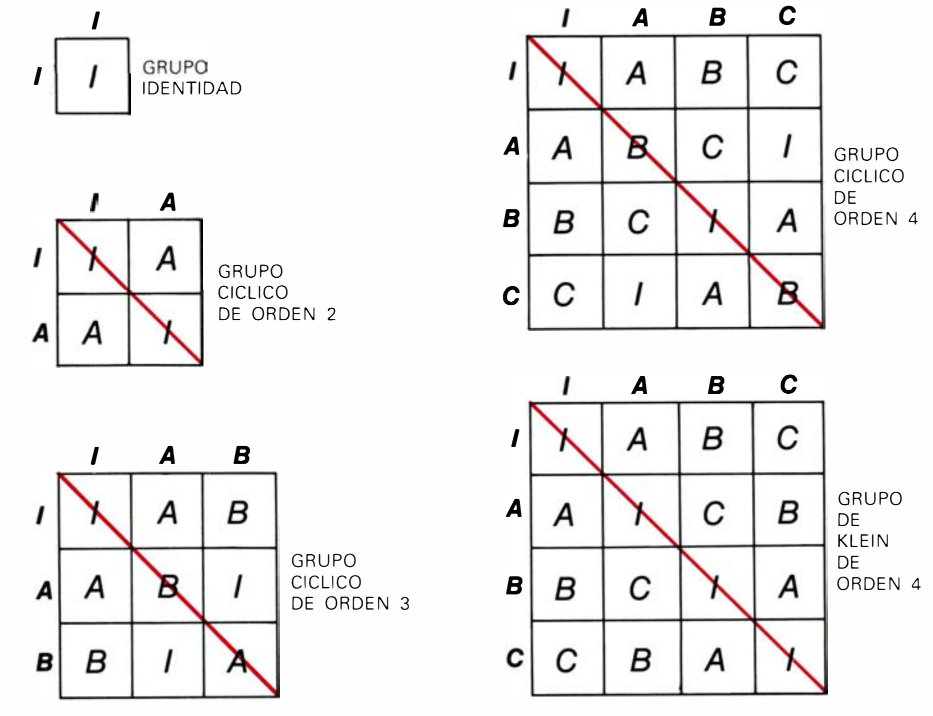
1. Axioma de clausura: El producto de cualquier par de operaciones es equivalente a una sola operación de las pertenecientes al conjunto.

2. Propiedad asociativa: Si el producto de cualesquiera dos operaciones se hace seguir de otra operación cualquiera, el resultado es el mismo que al hacer seguir la primera operación del producto de la segunda y la tercera.

3. Elemento neutro (o identidad): Existe precisamente una operación que no efectúa ninguna modificación; en nuestro caso, permanecer firmes (no hacer nada).

4. Existencia de inversos: Para cada operación existe una operación inversa tal que el resultado de efectuar primero la operación y luego la operación inversa es equivalente a efectuar la operación identidad. En nuestro ejemplo, "izquierda" y "derecha" son mutuamente inversas, mientras que "firmes" (elemento neutro) y "media vuelta" son inversas de sí mismas.

Cualquier colección de operaciones que satisfaga los cuatro axiomas anteriores formará un grupo. El grupo de cuatro órdenes que acabamos de ver se llama grupo cíclico de orden cuatro, pues equivale al grupo de las permutaciones circulares de cuatro objetos dispuestos en fila. (Al permutar cíclicamente un conjunto ordenador de elementos, el primer objeto pasa a ocupar la segunda posición, el segundo, a la tercera, y así sucesivamente, hasta el último elemento, que saltará al primer lugar.) Bauticemos 1, 2, 3 y 4 a los cuatro objetos, que imaginaremos alineados en orden numérico: 1234. La operación neutra (llamada también identidad), que denotamos I, deja intacto el orden relativo de los cuatro objetos. La operación A los redispone



GRUPO IDENTIDAD

	I
I	I

GRUPO CICLICO DE ORDEN 2

	I	A
I	I	A
A	A	I

GRUPO CICLICO DE ORDEN 3

	I	A	B
I	I	A	B
A	A	B	I
B	B	I	A

GRUPO CICLICO DE ORDEN 4

	I	A	B	C
I	I	A	B	C
A	A	B	C	I
B	B	C	I	A
C	C	I	A	B

GRUPO DE KLEIN DE ORDEN 4

	I	A	B	C
I	I	A	B	C
A	A	I	C	B
B	B	C	I	A
C	C	B	A	I

Los únicos grupos de órdenes 1 a 4

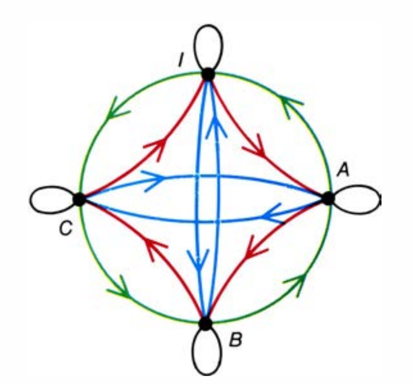
en la forma 4123, la B, en 3412, y la C, en el orden 2341. El grupo queda enteramente caracterizado por la tabla de "multiplicar" de la parte superior derecha de la ilustración de la página anterior. En cada casilla está expresada la operación equivalente a realizar primero la indicada en el extremo izquierdo de su fila, y seguidamente, la indicada en el encabezamiento de su columna. Si denotamos I, A, B y C las cuatro órdenes ("firmes", "izquierda", etcétera) respectivamente a como fueron enumeradas, se obtiene una tabla idéntica, lo que demuestra que el grupo cíclico de orden 4 y el grupo de cuatro órdenes son isomorfos, es decir, equivalentes.

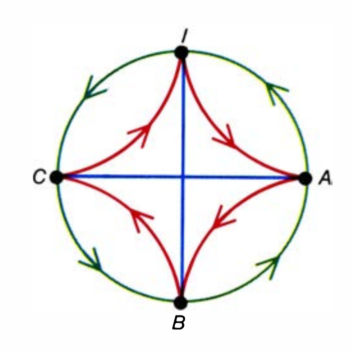
Observemos que la tabla del grupo cíclico de orden 4 es simétrica respecto de una de sus diagonales, como se muestra en color. Esta característica indica que el grupo obedece también la ley conmutativa, lo que significa que al multiplicar dos operaciones cualesquiera es indiferente cuál se ejecuta en primer lugar. Los grupos que exhiben esta propiedad se llaman grupos abelianos, en honor del matemático noruego Niels Henrik Abel. La colección de todas las permutaciones cíclicas de n objetos forma un grupo abeliano equivalente al grupo de giros que dejan invariable un polígono regular de n lados. (Se dice que un giro deja invariante a una figura cuando la posición final de ésta tras la rotación es indistinguible de la posición inicial). Así pues, también podemos tomar por modelo del grupo cíclico de orden 4 al grupo de giros de un cuadrado.

Existe exactamente un grupo de orden 1: el grupo trivial, formado por la operación neutra. Nada cuesta comprobar que esta sola operación cumple los cuatro requisitos de la definición de grupo. Por ejemplo, dejar intacta una cosa por dos veces sucesivas es lo mismo que dejarla intacta una vez, y por tanto, se verifica en este caso el axioma de clausura. No menos trivial es el único grupo de orden 2. Este grupo, cuya tabla vemos en la ilustración de la página opuesta, puede materializarse con dos operaciones a efectuar sobre una moneda: dejarla intacta (I), y volverla de la otra cara (A). El único grupo de orden 3 es cíclico y equivalente al conjunto de permutaciones circulares de tres objetos, y equivalente también al conjunto de giros que dejan invariante un triángulo equilátero. Hay exactamente dos grupos de orden 4: el cíclico, y otro, conocido por el nombre de grupo de Klein.

Es fácil materializar el grupo de

	I	A	B	C
I	I	A	B	C
A	A	B	C	I
B	B	C	I	A
C	C	I	A	B



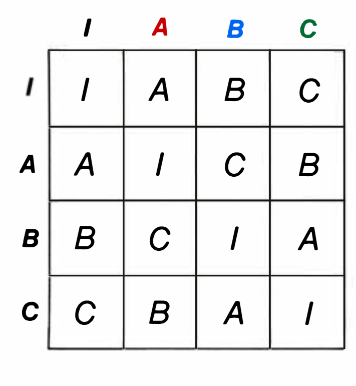


Grafo de Cayley del grupo cíclico de orden **4** *(***abajo, izquierda***), y su versión simplificada (***abajo, derecha***)*

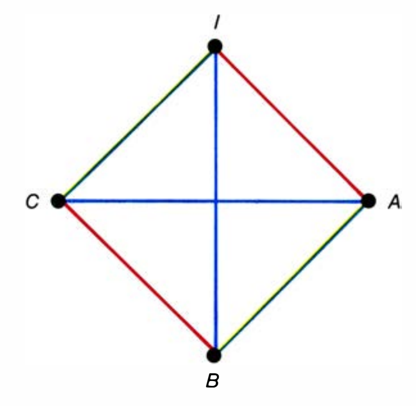
Klein colocando dos monedas, una junto a otra, y efectuando las siguientes operaciones: dejarlas quietas (I), volver de la otra cara solamente la moneda izquierda (A), ídem sólo la derecha, (B), y volver ambas a la vez (C). La tabla de este grupo, que vemos en la ilustración de la página contigua, revela que también este grupo es abeliano.

El ejemplo más sencillo de grupo no conmutativo resulta de tomar los seis movimientos o congruencias, que dejan fijo un triángulo equilátero, incluidas las simetrías, y que son: la identidad, el giro de 120 grados en sentido horario, el giro de 120 grados en sentido antihorario y las tres simetrías axiales, que equivalen a volver reverso por anverso alrededor de cada una de las tres alturas. Para comprobar que los elementos de este grupo no conmutan, recortemos en cartón un triángulo equilátero, y marcando sus vértices con letras, hagámoslo girar 120 grados en cualquier sentido, volviéndolo después de la otra cara alrededor de una altura. Ejecutemos luego estas operaciones en orden contrario, y comparemos los resultados. Identificando los vértices con tres objetos distintos se ve fácilmente que el grupo de seis elementos obtenido resulta equivalente al grupo de las permutaciones de tres objetos.

Para poner a prueba si verdaderamente ha comprendido estos grupos,



	I	A	B	C
I	I	A	B	C
A	A	I	C	B
B	B	C	I	A
C	C	B	A	I



Grafo de Cayley simplificado para un grupo de Klein

deténgase a examinar los tres modelos siguientes.

1. Con cuatro naipes apilados boca abajo se efectúan las siguientes operaciones: identidad (I), transposición de las dos cartas superiores del montón (A), transposición de las dos inferiores (B), y extracción de las dos intermedias, colocando después la inferior de las dos en lo más bajo, y la de encima, en lo más alto, (C).

2. Un billete de banco puede colocarse sobre la mesa en las posiciones anverso, reverso, lado derecho hacia arriba y lado derecho hacia abajo. Las operaciones son: identidad (I), giro de 180 grados (A), volverlo de la otra cara alrededor de su eje transversal (B) y, por último, volverlo alrededor del eje longitudinal, (C).

3. Cuando nos ponemos un calcetín podemos meternoslo en el pie derecho, o en el izquierdo, y el calcetín, del derecho o del revés. Las operaciones son: identidad (I), quitarnos el calcetín, darle la vuelta y ponérnoslo en el mismo pie (A), cambiamos de pie el calcetín sin darle la vuelta (B) y quitárnoslo, darle la vuelta, y meter en él el otro pie (C).

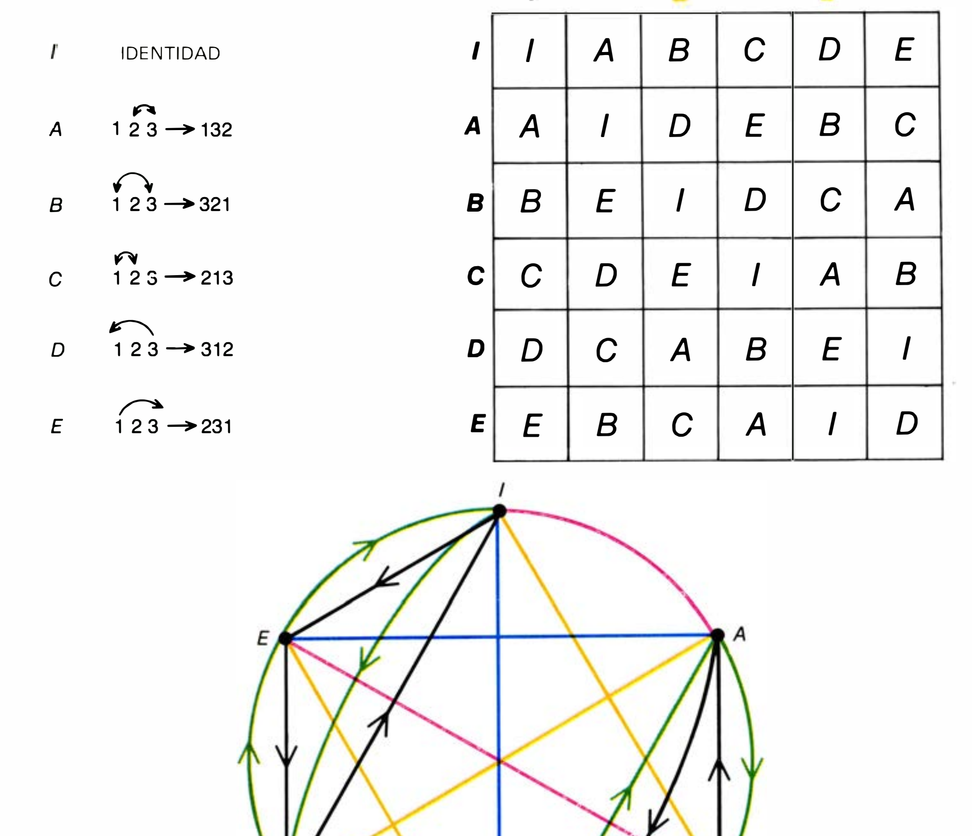
Se debe construir la tabla de multiplicar de cada uno de estos grupos, y determinar si el grupo es equivalente al grupo cíclico de orden 4 o al grupo de Klein. Daremos las respuestas el mes que viene.

La tabla de multiplicar de un grupo puede representarse gráficamente mediante un diagrama llamado grafo coloreado de Cayley, en recuerdo del matemático británico Arthur Cayley. Así, por ejemplo, el grafo inferior izquierdo de la ilustración superior de la página precedente es un grafo coloreado de Cayley, correspondiente al grupo cíclico de orden 4, cuya tabla vemos en la misma ilustración. Los cuatro vértices del grafo denotan las cuatro operaciones del grupo. Cada par de vértices está conectado por un par de líneas que van en sentidos opuestos, como muestran las puntas de flecha. Cada color corresponde a una operación, conforme a la clave mostrada sobre la tabla del grupo. Para ver la forma en que el grafo reproduce la información contenida en la tabla, fijémonos en la línea que va desde B hasta A. Para determinar su color se procede así: partimos de B, en el costado izquierdo de la tabla, y avanzamos por su fila hasta la casilla ocupada por A; tomamos entonces el color asignado a la letra (en este caso C) que encabeza la columna de la casilla. La misma regla determina el color de todas las demás líneas.

Cuando dos puntos del diagrama de Cayley están enlazados por líneas de distinto color, las operàciones representadas por los colores son mutuamente inversas. Si ambas líneas son de igual color, la operación asociada a tal color es inversa de sí misma. En este caso podemos simplificar el grafo sustituyendo las dos líneas orientadas, del mismo color, por una sola, desprovista de orientación. Por otra parte, la operación identidad está representada por bucles que enlazan cada punto consigo mismo, y por esta razón podemos darlos todos por sobreentendidos, ahorrándonos su trazado. Vemos la versión simplificada del grafo en la parte inferior derecha de la ilustración superior de la página precedente.

El grafo de Cayley simplificado correspondiente al grupo de orden 4 de Klein ocupa la parte baja de la página anterior, y en la ilustración de esta misma vemos otro, correspondiente al grupo no abeliano de permutaciones de tres objetos, que tiene orden 6. Para grupos de órdenes superiores ya no es práctico usar colores, siendo preferible rotular cada línea con el símbolo asociado a la operación que representa.

Sin duda es evidente que dado el grafo coloreado de un grupo cualquiera podemos construir su tabla de multiplicar. También la recíproca es verdadera. Empero, los grafos suelen ser ayuda valiosa porque suelen revelar propiedades no siempre fáciles de reconocer por la tabla de multiplicar. Por ejemplo, no es difícil ver que si en el grafo de orden 6 se omiten las líneas correspondientes a A, B y C, dejando únicamente las correspondientes a D y a E, se obtienen dos grafos desconectados. Cada uno de ellos podría corresponder a un grupo cíclico de orden 3, pero en realidad, solamente el conjunto de operaciones I, D, E forma un tal grupo, porque sólo este conjunto contiene la operación identidad. Toda colección de elementos de un grupo que por sí misma forme grupo recibe el nombre de subgrupo; la inspección del



	I	A	B	C	D	E
I	I	A	B	C	D	E
A	A	I	D	E	B	C
B	B	E	I	D	C	A
C	C	D	E	I	A	B
D	D	C	A	B	E	I
E	E	B	C	A	I	D

Grafo de Cayley simplificado asociado al grupo de permutaciones de orden 6

NOMBRE GRUPO	NUMERO DE ELEMENTOS	DESCUBRIDOR
M_{11}	$2^4 \times 3^2 \times 5 \times 11$	
M_{12}	$2^6 \times 3^3 \times 5 \times 11$	
M_{22}	$2^7 \times 3^2 \times 5 \times 7 \times 11$	Mathieu
M_{23}	$2^7 \times 3^2 \times 5 \times 7 \times 11 \times 23$	
M_{24}	$2^{10} \times 3^3 \times 5 \times 7 \times 11 \times 23$	
J_1	$2^3 \times 3 \times 5 \times 7 \times 11 \times 19$	Janko
J_2	$2^7 \times 3^3 \times 5^2 \times 7$	Hall, Wales
J_3	$2^7 \times 3^5 \times 5 \times 17 \times 19$	Higman, McKay
J_4	$2^{21} \times 3^3 \times 5 \times 7 \times 11^3 \times 23 \times 29 \times 31 \times 37 \times 43$	Benson, Conway, Janko, Norton, Parker, Thackray
HS	$2^9 \times 3^2 \times 5^3 \times 7 \times 11$	Higman, Sims
MC	$2^7 \times 3^6 \times 5^3 \times 7 \times 11$	McLaughlin
Sz	$2^{13} \times 3^7 \times 5^2 \times 7 \times 11 \times 13$	Suzuki
C_1	$2^{21} \times 3^9 \times 5^4 \times 7^2 \times 11 \times 13 \times 23$	
C_2	$2^{18} \times 3^6 \times 5^3 \times 7 \times 11 \times 23$	Conway
C_3	$2^{10} \times 3^7 \times 5^3 \times 7 \times 11 \times 23$	
He	$2^{10} \times 3^3 \times 5^2 \times 7^3 \times 17$	Held, Higman, McKay
F_{22}	$2^{17} \times 3^9 \times 5^2 \times 7 \times 11 \times 13$	
F_{23}	$2^{18} \times 3^{13} \times 5^2 \times 7 \times 11 \times 13 \times 17 \times 23$	Fischer
F_{24}	$2^{21} \times 3^{16} \times 5^2 \times 7^3 \times 11 \times 13 \times 17 \times 23 \times 29$	
Ly	$2^8 \times 3^7 \times 5^6 \times 7 \times 11 \times 31 \times 37 \times 67$	Lyons, Sims
O	$2^9 \times 3^4 \times 5 \times 7^3 \times 11 \times 19 \times 31$	O'Nan, Sims
R	$2_{14} \times 3^3 \times 5^3 \times 7 \times 13 \times 29$	Conway, Rudvalis, Wales
F_5	$2^{14} \times 3^6 \times 5^6 \times 7 \times 11 \times 19$	Conway, Fischer, Harada, Norton, Smith
F_3	$2^{15} \times 3^{10} \times 5^3 \times 7^2 \times 13 \times 19 \times 31$	Smith, Thompson
F_2	$2^{41} \times 3^{13} \times 5^6 \times 7^2 \times 11 \times 13 \times 17 \times 19 \times 23 \times 31 \times 47$	Fischer, Leon, Sims
F_1	$2^{46} \times 3^{20} \times 5^9 \times 7^6 \times 11^2 \times 13^3 \times 17 \times 19 \times 23 \times 29 \times 31 \times 41 \times 47 \times 59 \times 71$	Fischer, Griess

La lista de los 26 grupos finitos simples esporádicos

grafo coloreado ha revelado que el grupo cíclico de orden 3 constituye un subgrupo del grupo de permutaciones de orden 6.

Hasta aquí nos hemos referido solamente a grupos formados por número finito de operaciones o elementos. Pero hay también grupos infinitos, que se dividen en dos categorías: grupos discretos, que contienen una infinidad numerable de elementos, y grupos continuos, cuyos elementos forman una infinidad no numerable. Se dice que un conjunto infinito es numerable cuando sus elementos pueden ponerse uno por uno en correspondencia biunívoca con los enteros positivos 1, 2, 3,... Por tanto, los propios números enteros sirven de ejemplo de conjunto numerablemente infinito; mientras que el conjunto de puntos de la recta real sirve como ejemplo de conjunto infinito no numerable. De hecho, los enteros forman respecto de la adición un grupo abeliano discreto, siendo 0 el elemento identidad, y $-a$ el inverso (que ahora se llama opuesto) de a, cualquiera que sea éste.

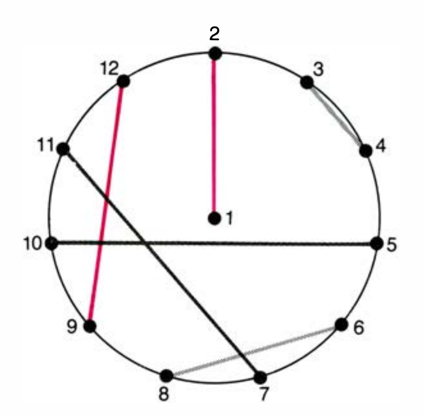
Por otra parte, los números reales forman un grupo continuo respecto de la adición, y si de ellos suprimimos el 0, también forman otro grupo continuo (¡distinto!) respecto de la multiplicación. (En este último caso, el elemento identidad es 1, y el inverso de a es $1/a$.) Los grupos continuos se llaman también grupos de Lie, en honor del matemático noruego Marius Sophus Lie. Ejemplos geométricos triviales de grupos de Lie son los grupos de simetrías de rotación de circunferencias (esferas, hiperesferas), pues la amplitud de los giros puede ser tan pequeña como se desee.

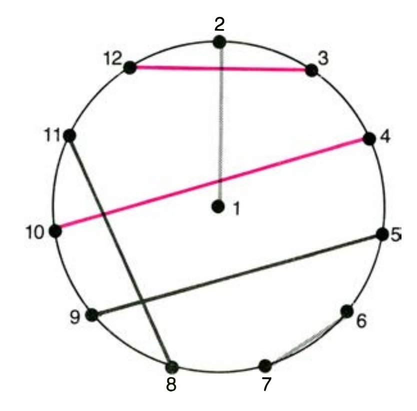
La noción de grupo es una de las concepciones con mayor capacidad de unificación de la matemática. Además, no sólo se presenta esta noción en todas las ramas de la matemática, sino que sus aplicaciones a las restantes ciencias son innumerables. Dondequiera se presenten simetrías aparecerán grupos. En teoría de la relatividad, las transformaciones de Lorentz forman un grupo de Lie basado en la rotación continua de un objeto de espacio-tiempo. Subyacentes a la estructura de los cristales hay grupos finitos, que, por otra parte, son indispensables en química, mecánica cuántica y física de partículas. El famoso método del octeto, que clasifica la familia de partículas subatómicas conocidas por hadrones, está basado en un grupo de Lie. Toda geometría puede ser definida como el estudio de las propiedades de una figura que permanecen invariables para un grupo de transformaciones.

A veces, incluso en matemática recreativa tiene interés la noción de grupo. Dado que todo grupo finito puede representarse mediante una colección de permutaciones ejecutadas sobre n objetos, poco puede sorprendernos hallar grupos en relación con barajaduras, malabarismos, campanología (repiques de campanas) y en rompecabezas y problemas de todo pelaje. En una sección publicada hace años (reproducida en mis *Nuevos Pasatiempos Matemáticos*, Alianza Editorial, Madrid) expliqué ya lo bien que encajan los grupos en la teoría de trenzados, y cómo, por consiguiente, están implícitos en numerosos trucos de ilusionismo con cuerdas y pañuelos.

En vista de la gran elegancia y utilidad de los grupos, es comprensible que los matemáticos hayan pretendido clasificarlos. Los grupos de Lie han podido serlo ya, pero hay grupos infinitos de otros tipos, refractarios a la clasificación. ¿Y qué podemos decir de los grupos finitos? Podría suponerse que habrían de ser más fáciles de clasificar que los grupos de Lie, pero no ha resultado ser así. Y esta difícil tarea es la que ahora dista poco de quedar rematada.

Todos los grupos finitos están compuestos de bloques constructivos, llamados grupos simples, de una forma que recuerda mucho la formación de combinaciones químicas a partir de elementos puros, la síntesis de proteínas a partir de aminoácidos, o la formación de números compuestos a partir de números primos. Se dice que





Dos soluciones al problema del campeonato de bridge, del mes pasado

un grupo es simple cuando no tiene otros subgrupos normales que él mismo y el subgrupo trivial, a saber: la identidad.

Recordemos que los subgrupos han sido definidos ya como aquellos subconjuntos del grupo que son grupos por sí mismos. Para explicar el significado de "normal" lo mejor puede ser esto: Consideremos un grupo *G*, y uno de sus subgrupos, *S*. Para cada uno de los elementos de *G* formemos todos los productos de *g* por cada uno de los elementos de *S;* el conjunto de productos así obtenidos se llama cogrupo izquierdo de *S*, y se denota *gS*. Análogamente se define el cogrupo derecho *Sg*. Sucede con frecuencia que los cogrupos *gS* y *Sg* correspondientes a un mismo elemento por distintos lados sean distintos. Cuando los cogrupos *gS* y *Sg* son siempre iguales, cualquiera que sea el elemento *g* que tomemos en *G*, se dice que el subgrupo *S* es normal. Así, por ejemplo, el grupo cíclico de orden 3 es normal en el grupo de permutaciones (también llamado grupo simétrico) de orden 6. Por consiguiente, el grupo simétrico de orden 6 no es simple. Los grupos simétricos son los bloques constituyentes de todos los grupos, y por tanto, para clasificar los grupos finitos, es necesario clasificar antes los grupos simples.

Casi todos los grupos simples pertenecen a familias formadas por infinidad de miembros. Las familias de que hablamos permiten sistemas de clasificación francamente satisfactorios, pues hay procedimientos para construir cada uno de sus miembros, es decir, cada uno de los grupos pertenecientes a ellas. Por ejemplo, los grupos de permutaciones cíclicas que sean de orden primo (que podemos materializar mediante rotaciones de polígonos regulares de número primo de lados) son todos grupos simples. Más aún, son los únicos grupos finitos simples y abelianos, y también, los únicos que son cíclicos. Un famoso resultado, conocido como teorema de Lagrange, afirma que el orden, o sea, el número de elementos de un subgrupo cualquiera tiene que ser divisor del orden del grupo donde está contenido. Como los números primos carecen de divisores (exceptuados 1 y ellos mismos), se deduce de este teorema que ningún grupo de orden primo contiene subgrupos no triviales (distintos de la identidad y de sí mismo). Y si un grupo carece de subgrupos no triviales no tendrá subgrupos normales; por ello, todo grupo de orden primo es simple.

Otra importante familia de grupos finitos simples está compuesta por los grupos alternados, cuyos prototipos son los grupos de todas las permutaciones de clase par efectuadas sobre *n* objetos, para todo entero *n* mayor que 4. Las permutaciones de clase par pueden ser realizadas en un número par de etapas, cada una consistente en intercalar entre sí dos de los *n* objetos. Por ejemplo, el grupo cíclico de orden 3 es también un grupo alternado, porque 231 se obtiene de 123 en dos pasos (transponer primero 1 y 2, y luego, transponer 1 y 3). Lo mismo puede decirse para cualesquiera otras dos de las tres permutaciones circulares de tres objetos. Exactamente la mitad de las permutaciones son de clase par, y dado que hay $n!$ maneras de permutar *n* objetos, cada grupo alternado tiene orden $n!/2$. Las permutaciones de clase impar no forman grupos, porque el resultado de efectuar una tras otra cualesquiera dos permutaciones impares es una permutación de clase par, y por tanto no se verifica el axioma de clausura.

Hay otras 16 familias infinitas de grupos simples, todas formadas por grupos no conmutativos y no cíclicos. Los órdenes de los grupos simples (excluidos los cíclicos) forman una sucesión infinita que comienza en 60, orden del grupo alternado asociado a cinco objetos. (Este grupo es equivalente al grupo de giros de un dodecaedro, o icosaedro, regulares). La sucesión comienza 60, 168, 360, 504, 660, 1092, 2448, 2520, 3420, 4080, 5616, 6048, 6072, 7800, 7920.... Insertando en esta sucesión infinita el 1 y todos los números primos, la sucesión resultante da los órdenes de todos los grupos simples de orden finito.

Desdichadamente, esta lista contiene un reducido número de grupos (que comienzan con el de orden 7920) imposibles de acomodar en ninguna familia infinita. Son los llamados grupos anómalos no abelianos, las ovejas negras que desafían toda clasificación. En matemática se los conoce por grupos simples esporádicos, y son extraordinariamente complicados. De existir una colección infinita de estos esporádicos, y de no existir ninguna pauta que permitiera organizarlos, la tarea de clasificar los grupos simples finitos no pasaría del acto fallido. Empero, hay razones que indican la inexistencia de esporádicos distintos de los ya localizados, que son 26.

La búsqueda de grupos simples esporádicos comenzó en el decenio de 1860, al ser descubiertos los cinco primeros por el matemático francés Emile Leonard Mathieu. El menor de todos ellos, denotado M_{11}, está formado por 7920 operaciones, y tiene por modelo cierto tipo de permutaciones con 11 objetos. Hubo de transcurrir un siglo antes de ser descubierto el sexto esporádico, de orden 175.560, hallado en 1965 por Zvonimir Janko, de la Universidad de Heidelberg. Tres años más tarde, John Horton Conway, de la Universidad de Cambridge, sorprendía a todos, dando a conocer tres esporádicos más. Su trabajo se fundaba en el retículo de Leech, otro matemático inglés. El retículo de Leech es un esquema de empaquetamiento denso de hiperesferas en el espacio de dimensión 24. (En el retículo de Leech cada hiperesfera es tangente a precisamente otras 196.500.)

Leech descubrió su retículo al trabajar sobre códigos de corrección de errores. Resulta que existe íntima conexión entre ciertos grupos esporádicos y los códigos empleados para reconstruir mensajes distorsionados por ruido. Dos de los grupos esporádicos de Mathieu, los M_{23} y M_{24}, están emparentados con el código Golay de corrección, que suele utilizarse con fines militares. En términos generales, un

buen código de corrección se funda en un subconjunto de esferas unitarias de un empaquetamiento denso, situadas dentro de él tan alejadas entre sí como sea posible.

Un divertido episodio en donde interviene M_{24} me ha sido recordado por John McKay, codescubridor de dos esporádicos, y especialista en ciencias de computación de la Concordia University de Montreal. En 1898, George Abram Miller, especialista norteamericano en teoría de grupos, publicó un artículo en *Messenger of Mathematics* (vol. 27, pp. 187-190) donde demostraba que M_{24} no existe. Como podría esperarse, esta publicación no figura en la recopilación de trabajos de Miller. MacKay opina que pudiera existir relación entre la malograda demostración de Miller y el hecho de que, aunque Eric Temple Bell cita en el índice de su *Development of Mathematics* (segunda edición, 1945) la página 445 junto al apellido Miller, en dicha página para nada se menciona a Miller. Sí hay en cambio un corrosivo ataque contra ciertos especialistas en grupos "apenas salidos del analfabetismo matemático" que se empeñan en descubrir nuevos grupos a base de "hozar con impertinencia" en categorías muy restrictivas definidas por ellos mismos, "al parecer, con el manifiesto propósito de dignificar sus cálculos dándoles un aire de ficticia generalidad".

Al principio de este año se había demostrado la existencia de un par de docenas de esporádicos, y de dos más, J_4 y F_1 existía la convicción de que eran auténticos. (Vemos en la pág. 111 una relación completa de estos 26 grupos esporádicos.) El grupo J_4, propuesto por Janko en 1975, fue al fin construido en febrero por David Benson, Conway, Simon P. Norton, Richard Parker y Jonathan Thackray, grupo de investigadores de Cambridge. La existencia de F_1 (¡el monstruo!), que es con mucho el mayor de todos los esporádicos, fue conjeturada independientemente por Griess y por Berndt Fischer, de la Universidad de Bielefeld, en 1973, y como ya he comentado, fue construido en enero de este año por Griess. Algunos otros esporádicos mucho menores, cuya construcción exigió largos cálculos con ayuda de ordenador, están inmersos de tal forma en F_1 que su existencia resulta casi trivial una vez conocida la de F_1. Y más todavía: para asombro de todos, la construcción de F_1 fue llevada a cabo por Griess enteramente a mano. Se dice que F_1 está basado en un grupo de congruencias de rotación ¡en un espacio de dimensión 196.883!

¿Es completa la lista de esporádicos conocidos, con sus 26 nombres? Casi todos los especialistas opinan que sí, pero la tarea de demostrar tal conjetura pudiera ser formidable. De hecho, es probable que para publicar una demostración completa, definitiva, hicieran falta más de 10.000 páginas impresas. Empero, es preciso reconocer que las demostraciones de teoría de grupos propenden hacia longitudes colosales. Una muy famosa, de John Thomson y Walter Feit, que entre otras consiguió probar la conjetura de William Burnside, a saber, que todos los grupos finitos simples no conmutativos son de orden par, ocupaba más de 250 páginas: todo un ejemplar de *The Pacific Journal of Mathematics* (vol. 13, págs. 775-1029; 1963).

Daniel Gorenstein, de la Rutgers University, esbozó en 1972 un programa de 16 pasos encaminado a completar la clasificación de los grupos simples finitos. Esta guía hacia la demostración final fue pronto mejorada y muy "acelerada" por Michael Aschenbacher, del Instituto de Tecnología de California. Ambos se cuentan entre los máximos especialistas en grupos de todo el mundo. (En enero, Aschenbacher ganó el muy preciado Premio Cole, en álgebra.) En mayo de 1977, Gorenstein confesaba al *New York Times* que había estado trabajando en el problema de clasificación cinco horas diarias, siete días a la semana, 52 semanas al año, desde 1959. En aquella ocasión declaraba: "Quiero resolver este problema porque sí, porque deseo resolverlo, no porque vaya a beneficiar a la humanidad." Al igual que otros muchos especialistas, Gorenstein está convencido de que no se descubrirán nuevos grupos esporádicos, y de que faltan solamente uno o dos años para conseguir una demostración de que la lista de 26 grupos está completa.

Desde luego, no hay forma de predecir si se podrán o no encontrar aplicaciones prácticas a resultados matemáticos cuyo descubrimiento estuvo motivado por consideraciones de orden teórico. Lo que sí sabemos es que los grupos ocupan el corazón mismo de la estructura del universo. La naturaleza parece preferir más bien grupos pequeños y sin complicaciones, pero ello podría ser ilusión nuestra, por ser más fácil encontrar aplicación a los grupos pequeños, particularmente en un mundo limitado a las tres dimensiones espaciales. ¿Podrá alguien garantizar que no habrá una remota fecha —si es que la humanidad sobrevive— en que hasta el monstruo encuentre alguna aplicación importante, por inconcebible que pueda parecernos hoy?

La ilustración de la página opuesta presenta dos soluciones del problema del mes pasado, consistente en preparar un campeonato de bridge para 12 jugadores, de forma que se enfrentasen diariamente en tres meses, durante 11 días, siendo cada jugador pareja de cada uno de los demás exactamente una vez, y contrincante, precisamente dos. La distribución para el primer día está dada por un disco donde cada pareja está conectada por una línea de color, y las mesas están representadas mediante líneas de colores a juego. Al girar el disco en el sentido de las agujas del reloj, un paso cada vez, se van generando cíclicamente las disposiciones correspondientes a los 10 días restantes.

La sección de marzo, dedicada a las damas, daba cuenta de una posición final para una partida bloqueada en 24 jugadas, la más corta posible si no se producen capturas. Dije entonces que se creía ser única tal disposición. Varios lectores han demostrado que no es así, y han hallado todas las posibles soluciones finales, que en total son 28. Tan sólo dos de ellas son simétricas (contando la que yo dí), y en 16 de ellas, de no ser el turno de las negras, las blancas podrían todavía hacer un movimiento más.

He recibido muchas variantes del "problema imposible", ideadas para garantizar la unicidad de la solución sin necesidad de imponer cotas superiores al par de números. Una de las más elegantes se debe a Barry Wolk, de la Universidad de Manitoba. Lo mismo que en la primera versión, S y P se refieren a dos números enteros desconocidos, ambos mayores que 1. S conoce solamente la suma de los números, y P conoce únicamente su producto.

S: "No veo modo de que puedas determinar mi suma".

P: (tras una pausa adecuada): "Lo que has dicho no me ha servido de nada. Sigo sin conocer tu suma".

S (tras pensar un rato): "Pues ahora ya sé tu producto".

¿De qué números se trata? Suponiendo que la conjetura de Goldberg sea verdadera, puede demostrarse que han de ser 5 y 6.

Taller y laboratorio

De los misterios del arco iris; en particular, sobre sus bellos arcos supernumerarios

Jearl Walker

En la sección de septiembre de 1977 les describí cómo pueden observarse los arcos iris comunes de primero y de segundo orden mediante un experimento sencillo. No les hablé entonces de las bandas de color suplementarias, llamadas arcos supernumerarios, que a veces acompañan a los arcos iris naturales. Este mes voy a describirles algunos experimentos que posibilitan la investigación de tales arcos supernumerarios.

El arco iris natural más corriente es el arco iris primario, o de primer orden, llamado así porque precisa de una reflexión de los rayos luminosos en el interior de las gotas de lluvia. Los arcos iris secundarios, o de segundo orden, más infrecuentes, precisan de dos reflexiones internas. Los arcos iris de órdenes superiores, en los que intervienen más de dos reflexiones internas, resultan probablemente demasiado tenues para que puedan verse en el cielo, pero sí pueden observarse en los experimentos que les describí en 1977. Mis observaciones respondían a cuestiones que me planteara mi abuela a propósito de un arco iris doble que decía haber visto. Las preguntas eran simples, pero las respuestas no.

En aquel primer examen quedaron sin contestar dos de las preguntas que me formuló mi abuela después de que yo hubiese comparado los colores del arco iris con los colores que proyecta un prisma de vidrio que se mantiene bajo la luz solar. Es evidente que mi respuesta excesivamente expeditiva no la dejó muy satisfecha. Una de las preguntas era: Si la separación de colores es en un arco iris igual que en un prisma, ¿por qué aparecen bandas de color suplementarias (habitualmente purpúreas) junto a los colores esperados? Las bandas adicionales se encuentran justo debajo del arco iris primario y (menos frecuentemente) justo encima del secundario. Los prismas no producen estas bandas de color suplementarias. La segunda pregunta era: ¿Cómo es posible que algunos arcos iris sean blancos? Volvemos a encontrarnos aquí con que no puede ser responsable de ellos una separación prismática de la luz.

Las bandas de color suplementarias son los arcos supernumerarios, nombre que implica que no debían presentarse. Lo cierto es que, si hemos de creer que las gotas de agua separan la luz blanca en sus componentes a la manera de un prisma, estos arcos supernumerarios resultan inesperados. Los experimentos que aclaran este fenómeno amplían los que entonces describí, pero ahora se incluyen los efectos de la interferencia ondulatoria de la luz que se dispersa desde una pequeña gota de agua. Estos experimentos, de fácil realización, revelan la belleza de la figura óptica de la que forman parte los colores del arco iris común.

En mi primera colaboración expuse cómo puede suspenderse una gotita de agua de un alambre que se mantiene vertical y después iluminarse con la luz blanca de un proyector. De todos los rayos luminosos que llegan a la gota, uno de ellos (el llamado rayo cartesiano en honor de René Descartes, que fue el primero en determinar su naturaleza) penetrará en la gota precisamente por el punto concreto que le permita emerger de ella para formar el arco iris primario. Los rayos que incidan sobre la gota en otros puntos no contribuirán a este arco iris.

En el montaje experimental que figura en la ilustración de la página 116, el rayo cartesiano que contribuye al arco iris primario penetra en la gota por el lado opuesto al observador. Después de refractarse en la gotita, este rayo se refleja una vez en la superficie interior y después sale de la gotita por el lado del observador. En la ilustración de la parte superior de la página 117, se representa la trayectoria del rayo vista desde arriba. Si se apunta adecuadamente el haz luminoso incidente (tarea que puede requerir un poco de paciencia), el rayo atravesará una sección transversal circular y horizontal de la gota y, cuando el observador tenga la cabeza en la posición correcta para interceptar el rayo emergente, se verán los colores del arco iris.

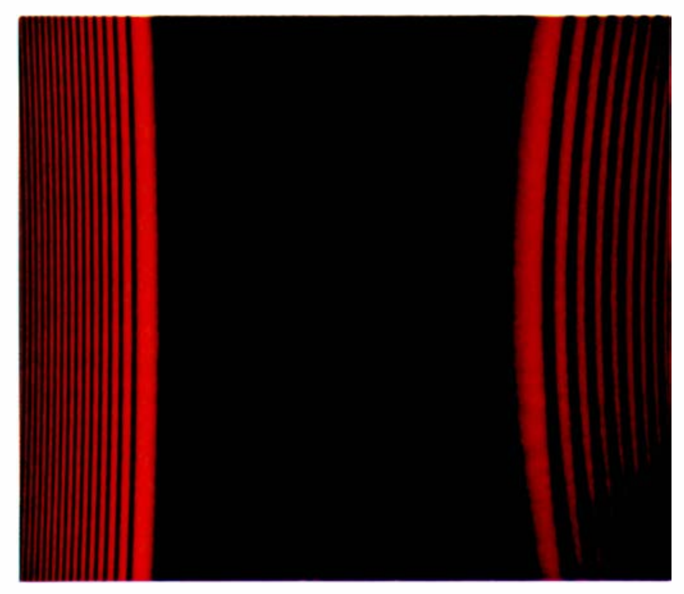
El rayo cartesiano correspondiente al arco iris primario se desvía unos 138 grados de su dirección de propagación inicial. Los demás rayos que penetran en la gotita y que se reflejan una vez en la superficie interior se desvían ángulos mayores, hasta los 180 grados que corresponden al rayo que entra en la gota por el centro y que es devuelto directamente a la fuente luminosa. Ningún rayo que se refleje una sola vez dentro de la gotita puede desviarse un ángulo menor que el del rayo cartesiano.

Si se ilumina la gotita con luz blanca, los colores componentes se refractan de forma ligeramente diferente y salen bajo ángulos distintos. El azul se refracta un poco más que el rojo, por lo que el rayo cartesiano azul gira respecto a su dirección de propagación original ligeramente más que el rojo. Los colores intermedios se desviarán en un rango cuantitativo intermedio.

Si se contempla la gotita bajo un ángulo próximo al ángulo del arco iris, los colores diferentes emergerán bajo ángulos diferentes y, en consecuencia, podrán distinguirse. El ángulo de desviación, medido respecto a la dirección inicial de la luz, es de unos 138 grados para el rojo y de 139 para el azul. Si nuestra gotita fuese una entre muchas que cayesen del cielo bajo la luz del Sol, un observador vería un arco de colores en el firmamento, es decir, el arco iris natural primario. En el caso de una gotita de agua suspendida de un alambre no existe tal arco, ya que se trata de una gotita aislada, pero sí pueden verse los colores sin dificultad en la cara de la gotita (la enfrentada a la fuente luminosa) por donde emergen los rayos.

Supongamos que la luz blanca se

Arco iris con arcos supernumerarios purpúreos fotografiado en California por Alan L. Falk



Figuras de interferencia de los dos primeros órdenes

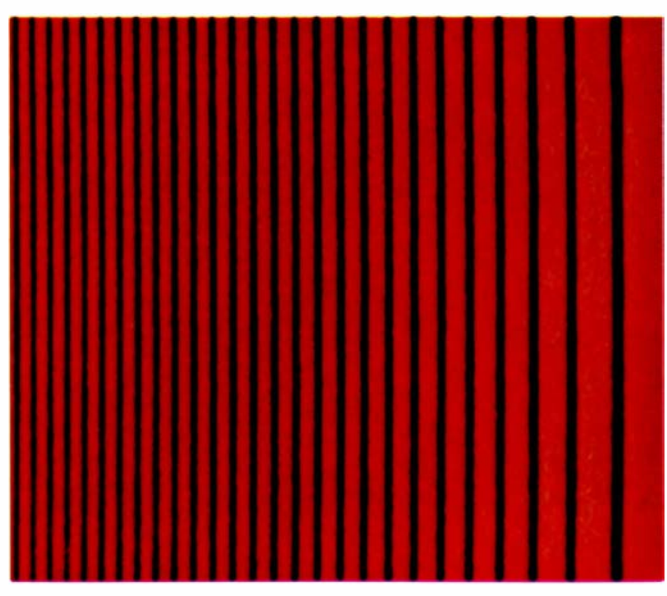


Figura de interferencia del arco iris primario

sustituye por luz monocromática procedente de una fuente tal como un láser de helio-neón, que emite a una longitud de onda de 632,8 nanómetros. Cabría esperar aquí que de la gotita sólo emergiese la componente roja del arco iris. Sorprendentemente, lo que se ve en realidad es una serie de bandas rojas verticales y no una banda roja única. La banda más ancha y más luminosa se encuentra en la posición angular que ocupa el rojo en los colores del arco iris producido por luz blanca. Las restantes bandas rojas, más débiles y más estrechas, se extienden en una serie que empieza en la banda más ancha y sigue hacia ángulos de desviación mayores. En esta serie de bandas adicionales se encierra la clave de la naturaleza de los arcos supernumerarios.

Sue Foroughi, de Council Bluffs, Iowa, ha investigado la figura de bandas rojas formada por una gotita de agua suspendida e iluminada por un láser de helio-neón. Mi diseño experimental lo ha perfeccionado sustituyendo el alambre por una bureta, de forma que ajustando cuidadosamente la espita de ésta fuera posible formar una gotita de agua en el extremo inferior del aparato. Seguidamente colocó la gotita en el centro de un espectrómetro con el fin de servirse del anteojo del aparato para observar la figura. (El efecto del anteojo es situar al observador a gran distancia de la gotita. Sin él, el ojo enfocaría la luz dispersada para dar en la retina una imagen de la gotita y no del espectro).

En especial, Foroughi se ha interesado por la relación entre tamaño de la gotita y número de bandas rojas que aparecen en la figura por grado. Sobre una gotita recién formada en el extremo de la bureta contó la densidad angular de bandas rojas desplazando el anteojo un ángulo de un grado. Tuvo que dar un valor aproximado del diámetro de la gotita. A continuación, esperó un tiempo para que la evaporación redujese el tamaño de la gotita y repitió el proceso.

La figura producida por una gotita de unos cuatro milímetros de diámetro inicial resultó tener una densidad de 16 bandas por grado en las cercanías de la banda principal del arco iris. Dos horas más tarde, el diámetro se había reducido a unos 2,5 milímetros y la densidad había descendido a ocho bandas. Foroughi halló que a medida que la gotita se encogía, aumentaba la anchura de las bandas y disminuía por tanto el número de éstas por grado. Resultado que nos da una pista sobre la naturaleza de los arco iris blancos por los que preguntaba mi abuela.

En mi montaje experimental, el rayo que permite al observador ver el arco iris secundario penetra en la gotita por el lado del observador. Se trata de un rayo cartesiano que se refracta en la gotita, se refleja dos veces en su superficie interior y emerge hacia el observador. Como antes, los distintos colores se refractan de forma suavemente distinta y así emergen bajo ángulos ligeramente diferentes. Un observador que se mueva entre los límites de los ángulos de estos colores verá el espectro que en el medio natural produce el arco iris secundario.

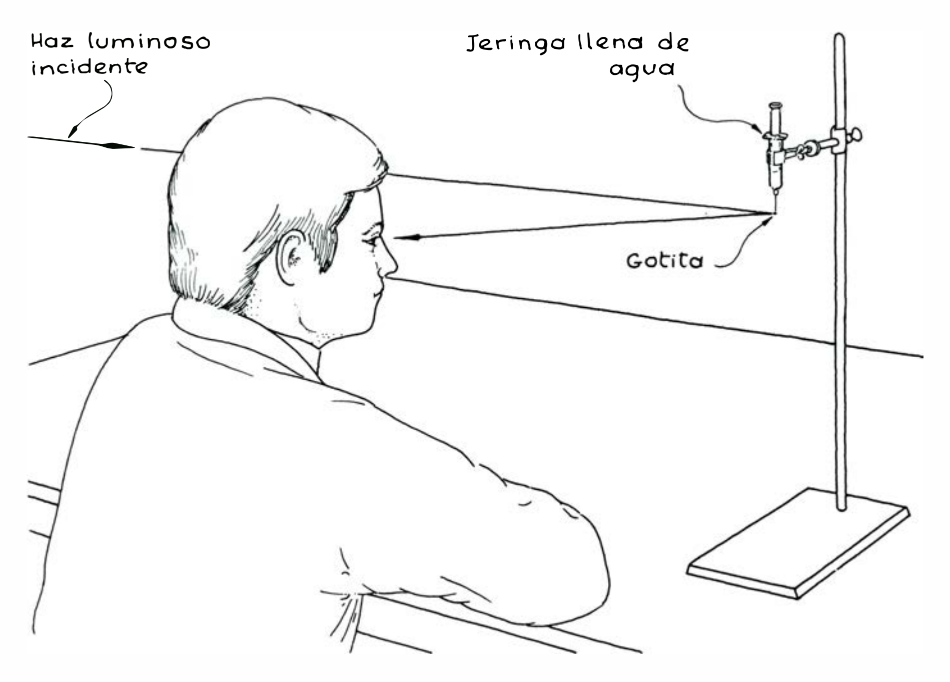
En el montaje experimental de la gotita suspendida, los colores aparecen en la cara de la gotita situada del lado de la fuente luminosa. El rayo cartesiano rojo emerge bajo un ángulo de unos 130 grados respecto a su dirección inicial de propagación, el rayo cartesiano azul a unos 127 grados. A estos ángulos voy a llamarlos ángulos de desviación del arco iris secundario.

Si la gota se ilumina totalmente, un observador puede recorrer los ángulos del arco iris primario y después los del secundario moviéndose desde un ángulo de desviación de unos 140 grados hasta un ángulo de unos 126 grados. La secuencia de colores irá del azul al rojo y seguidamente, tras una separación, del rojo al azul, ocupando los colores secundarios un espacio angular mayor. Entre ambas secuencias de colores la gotita estará relativamente oscura.

Esta banda oscura, que aparece también entre las franjas de los arcos iris naturales, se llama banda de Alejandro [*véase la ilustración inferior de la página siguiente*]. Su oscuridad es consecuencia de la relativa falta de rayos luminosos que emerjan de una gotita de agua sobre ese espacio angular. En efecto, todos los rayos que se reflejen una vez dentro de las gotitas deberán emerger bajo ángulos de desviación de unos 138 grados o más; todos los rayos que se reflejen dos veces dentro de la gotita emergerán bajo ángulos de desviación de unos 130 grados o más. Por esta causa el espacio intermedio de unos ocho grados queda relativamente oscuro.

Cuando una fuente de luz blanca se sustituye por un láser de helio-neón, el arco iris secundario completo es sustituido por otra distribución de bandas rojas. La banda más luminosa y ancha vuelve a encontrarse en la posición ocupada por el rojo en el arco iris completo y las bandas verticales se extienden desde la posición angular del rojo normal hacia ángulos de desviación más pequeños, estrechándose progresivamente a medida que aumenta su distancia angular a la banda más luminosa.

Este tipo de figura es característico y propio de los arcos iris de los dos primeros órdenes. Los órdenes adicionales exhiben cada uno su propia figura al iluminar una gota con luz láser monocromática. Por ejemplo, el arco iris de quinto orden resulta de cinco reflexiones internas de la luz y, aunque



Montaje de laboratorio para investigar el arco iris

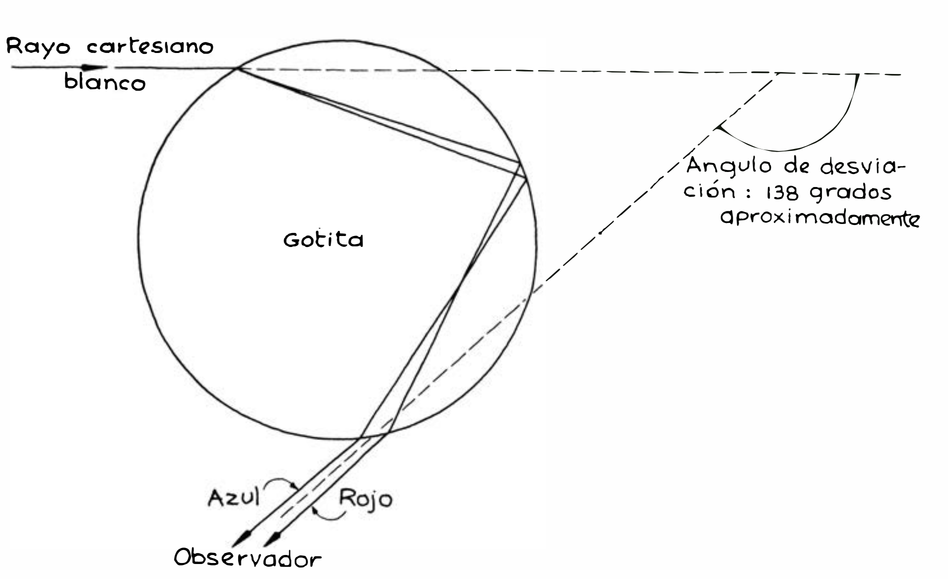
sea excesivamente tenue para que pueda verse en el cielo, sus colores son visibles en una gotita suspendida si se apunta con cuidado el haz de luz incidente.

El rayo cartesiano de quinto orden penetra en la gotita por el lado enfrentado al observador, se refleja cinco veces en el interior y emerge por la zona de la gotita enfrentada a la fuente luminosa. El rojo emerge bajo un ángulo de unos 127 grados con la dirección inicial de la luz, el azul bajo unos 134 grados. Cuando la luz blanca se sustituye por el láser, el arco iris de quinto orden da por resultado una distribución de bandas rojas, la más ancha y luminosa de las cuales está a unos 127 grados. Este espectro es más tenue que los espectros de los dos primeros órdenes. En general, cuanto mayor sea el orden de un arco iris, tanto más tenues y anchas serán las bandas principales de su espectro. Lo mismo ocurre en el caso de los arcos iris naturales: cuanto mayor es su orden, tanto más tenues y anchos resultan.

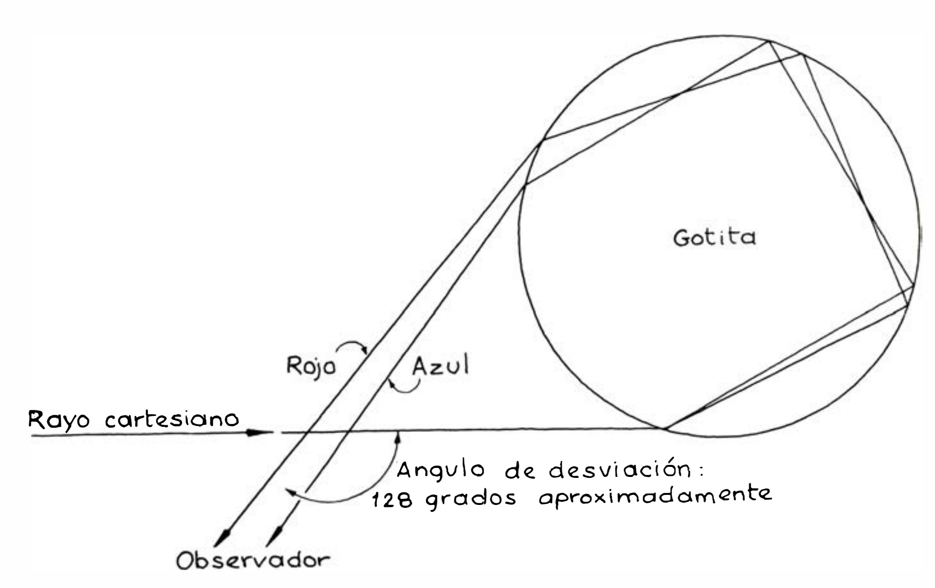
Cuando una gotita de agua se ilumina con luz blanca, cada uno de los colores produce su propia figura de bandas luminosas y oscuras en las cercanías de los ángulos del arco iris. Un arco iris natural se compone de las bandas principales (más anchas y brillantes) de cada color. Aunque en él también existan las bandas más estrechas y tenues del espectro de cada uno de los colores, éstas son habitualmente imperceptibles. En ocasiones, no obstante, las bandas más tenues se solapan con precisa corrección para hacerse visibles. Estas bandas son los arcos supernumerarios, que en el dispositivo experimental y en la naturaleza aparecen en el lado del azul del arco iris.

Cuando repetí algunos de los experimentos de Foroughi y algunos de los primeros míos, dispuse una gotita de agua colgando de la aguja de una jeringa. Para ello aplané la punta amolándola de modo que, con la jeringa apuntada hacia abajo, se pudiese formar una gotita que colgase convenientemente al empujar el émbolo. La óptica de un arco iris se hace mucho más complicada de lo descrito si los rayos luminosos no atraviesan la gotita de agua por una sección transversal circular y de aquí que sea deseable conseguir una gotita que cuelgue vertical y tenga una sección transversal circular y horizontal.

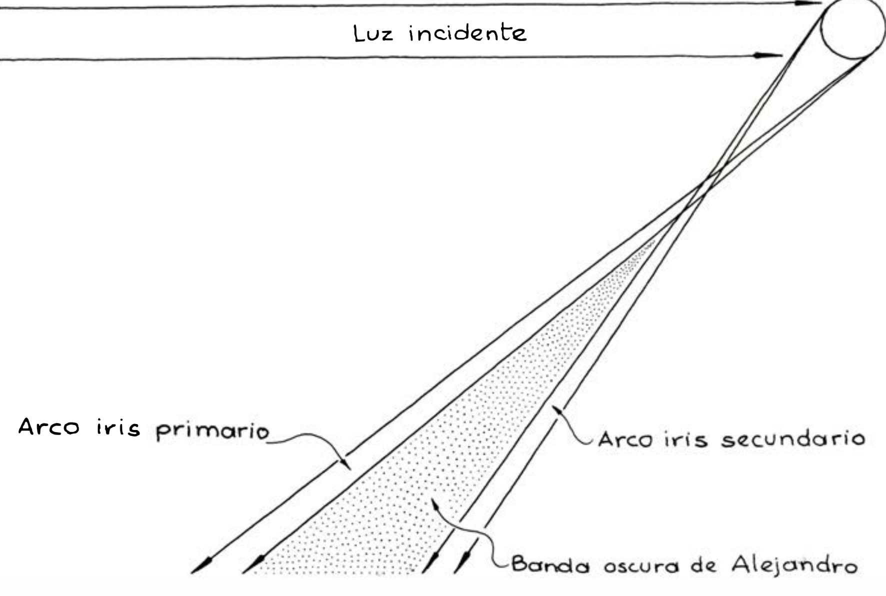
Coloqué la jeringa en posición utilizando una pinza para tubos de ensayo y



Dispersión de la luz en el arco iris primario

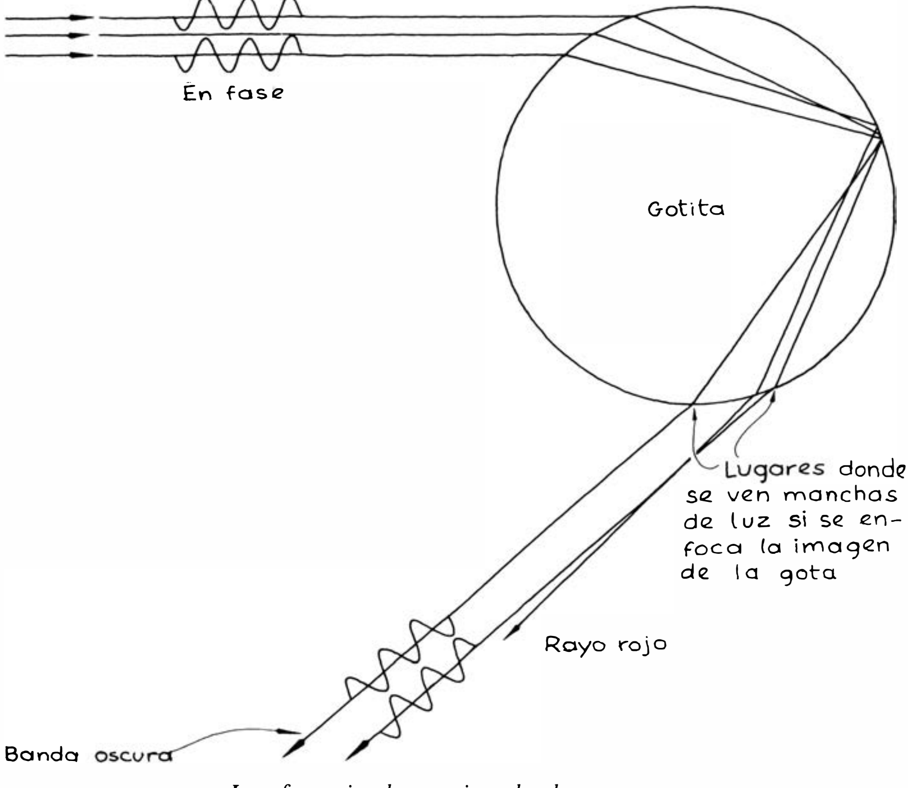


Cómo se dispersa la luz en el arco iris secundario

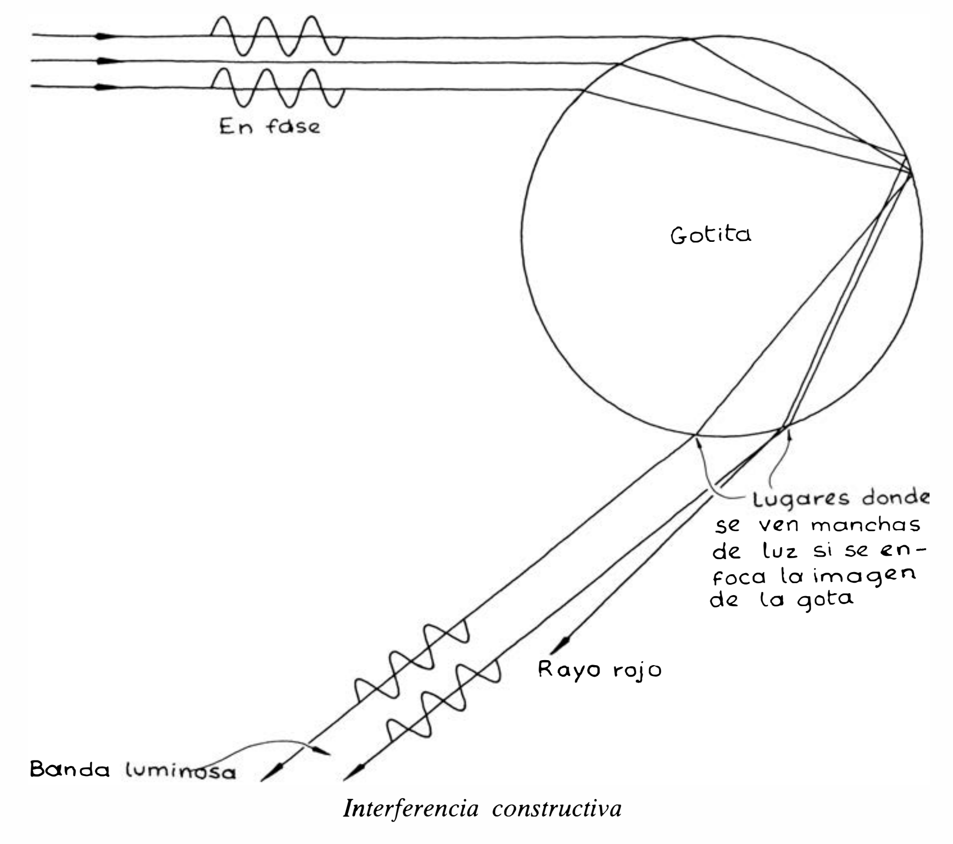


Circunstancias responsables de la banda oscura de Alejandro

un soporte de laboratorio, asentándola bien dentro de la pinza con un trozo de caucho esponjoso. Apretando poco a poco el émbolo, podía formar entonces una gota pequeña o una de gran tamaño en la punta roma de la aguja. Las gotas recibieron la iluminación de un láser de helio-neón de 35 miliwatt. A quien repita este ensayo le sugiero que emplee un láser menos potente o que coloque filtros en el haz con el fin de atenuarlo, ya que la luz reflejada en la aguja e incluso en la gota resultaba peligrosa e incómodamente brillante.



Interferencia destructiva de dos rayos



Interferencia constructiva

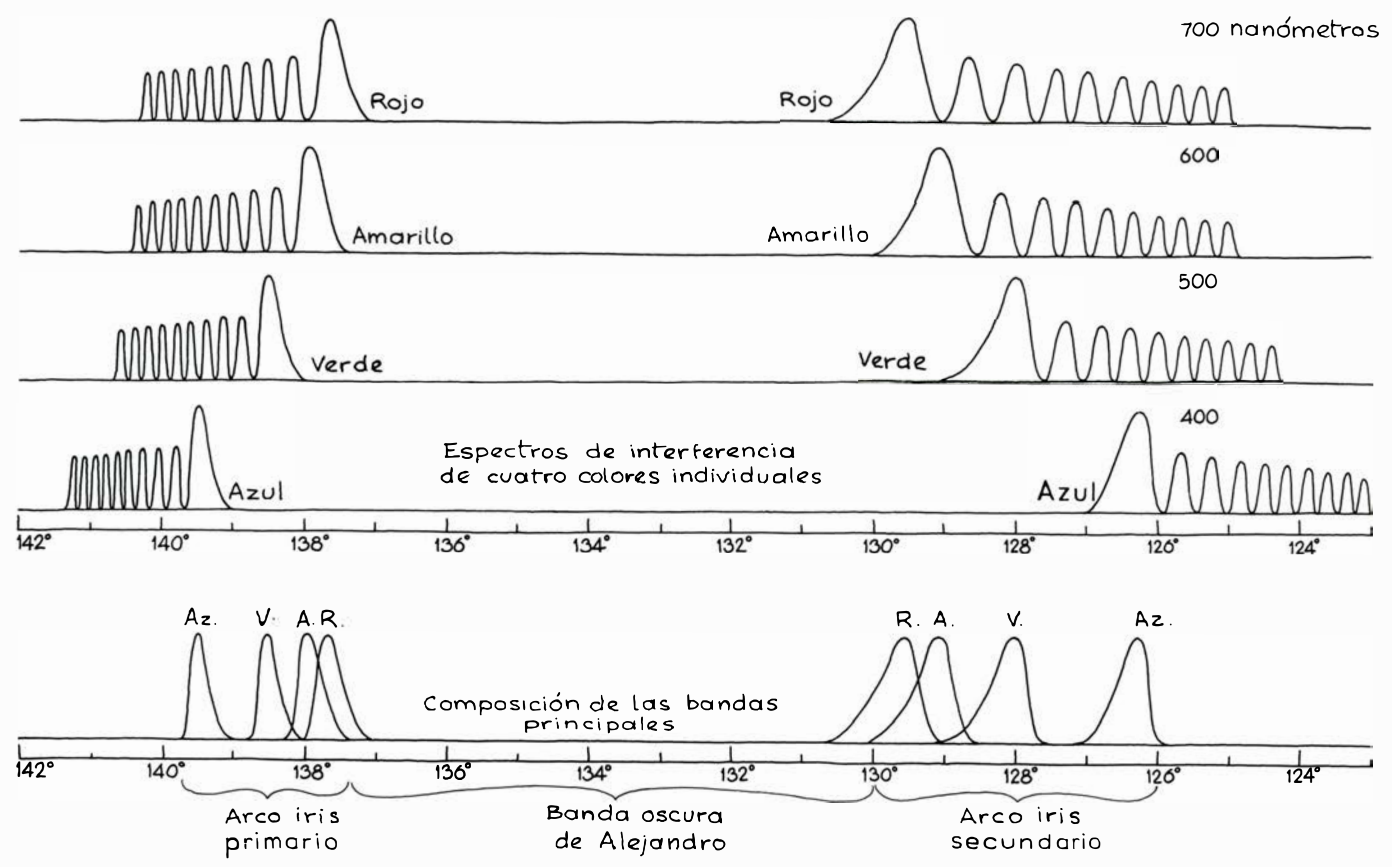
Si el láser es de luz polarizada, la polarización debe ser vertical para que la gota disperse la luz de la forma conveniente.

Regulando la iluminación sobre la gotita conseguí la figura tanto del arco iris primario como del secundario. Así, iluminando la zona de la gota más próxima a mí se creaba la figura del arco iris secundario e iluminando la otra zona se creaba el primario. Dado que el haz era más estrecho que las gotitas, al menos en el caso de las gotas mayores, no me era posible conseguir ambas figuras a la vez; sin embargo, al intercalar una lente en el haz, éste se abrió lo bastante para crear las dos simultáneamente.

Como mi láser era muy brillante las figuras se proyectaban en las paredes de la habitación con intensidad considerable. Además pude ponerlas de manifiesto sin dificultad sosteniendo una hoja de papel blanco cerca de la gota y examinarlas por el visor de mi cámara. Iluminando adecuadamente la zona más afectada de la gotita, pude ver los espectros parcialmente solapados de los arcos iris de primer y quinto orden.

La mejor forma de alinear la luz con la gotita es utilizando la sombra de ésta sobre una pared. Para ello, mientras contemplaba la sombra de la gotita manipulé el láser hasta centrar el haz en ella. Con este procedimiento, si deseaba iluminar uno de los lados de la gotita más que el otro, bastaba con desplazar el láser lo conveniente. A veces colocaba una tarjeta negra en el haz para obstruir la luz incidente sobre uno de los lados de la gotita. También empleé esta tarjeta para evitar que la luz del láser diese sobre la aguja, que de otro modo producía un fulgor deslumbrante.

Con mi cámara de 35 milímetros, cargada con película de 64 ASA y desprovista de objetivo, fotografié los espectros con una exposición de 1/125 de segundo. Para fotografiar los arcos iris de los dos primeros órdenes situé la cámara y el trípode lo más cerca posible de la gotita de agua suspendida sin interceptar el haz del láser con la cámara. Alejando la cámara de la gotita pude ampliar la imagen sobre la película. Esta operación ponía más de manifiesto sus detalles, pero acrecentaba los efectos adversos de las vibraciones de la gotita. Las vibraciones se debían a las corrientes de aire y a las vibraciones del edificio, que sacudían la gota y destruían momentáneamente la sección transversal circular perfecta



Figuras de interferencia que forman los arcos iris de los dos primeros órdenes

atravesada por los rayos luminosos, con lo que el espectro tremolaba. Reduje estas vibraciones a un mínimo afianzando el soporte con bloques de hormigón. Los efectos en cuestión podían rebajarse trasladando el experimento al sótano o montando la instalación sobre una plataforma de tipo antivibratorio como las que se usan para hacer hologramas.

Cuando quise fotografiar la gotita de agua, pero no las bandas rojas, monté en la cámara su objetivo normal (50 milímetros y f/1,4) y varias lentes de aproximación con el fin de enfocar los rayos dispersados por la gotita. La exposición adecuada hube de determinarla por tanteo y para asegurarme unas cuantas fotos de buena calidad accioné la cámara bajo muchas exposiciones.

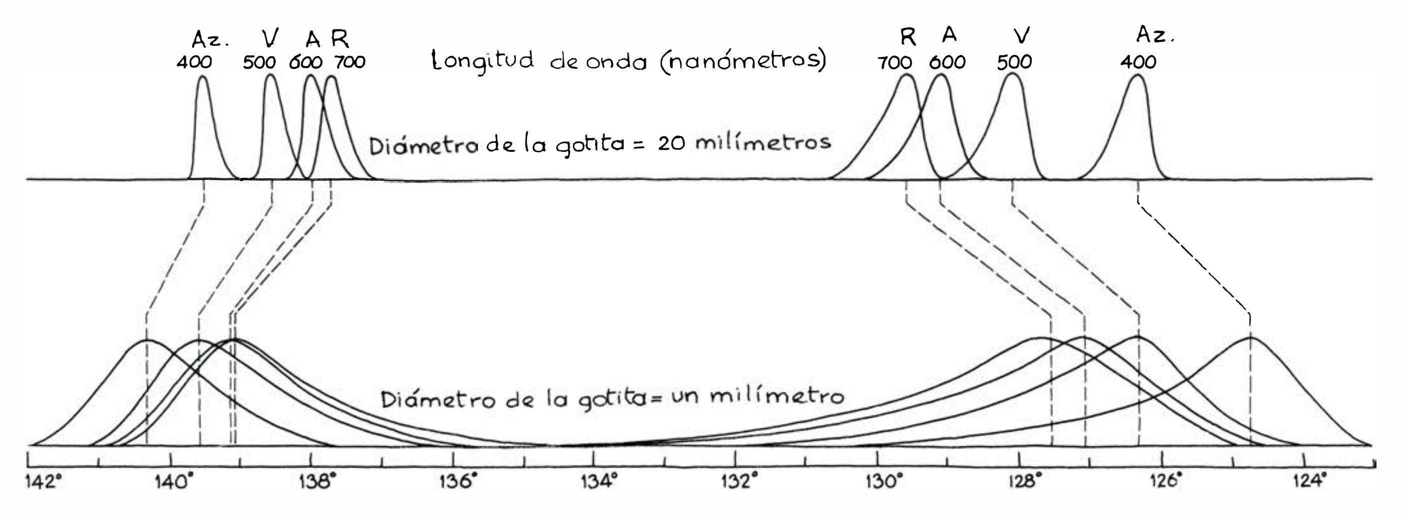
En lugar de esperar a que cada gotita encogiese por evaporación, observé las figuras de los arcos iris de los dos primeros órdenes mientras empujaba lentamente el émbolo de la jeringa. A medida que la gotita aumentaba de tamaño, pude ver cómo se estrechaban las bandas del espectro. Noté además otro efecto y es que las posiciones angulares de las bandas principales (la más luminosa y ancha de cada espectro) se desplazaban, de forma que a la vez que crecía la gotita las bandas principales de ambos órdenes se acercaban una a la otra, estrechándose así la zona oscura intermedia. Si la gota es grande (de unos cuatro milímetros de diámetro, más o menos), las figuras están relativamente próximas y se componen de bandas relativamente estrechas. Cuando la gota se encoge, las figuras se desplazan alejándose entre sí y las bandas se ensanchan.

¿Por qué una fuente de luz monocromática produce una distribución de bandas luminosas y oscuras en la proximidad de la posición angular normal del arco iris? Está claro que las gotas no separan los colores de la misma forma que lo haría un prisma. Thomas Young demostró en 1803 que las ondas luminosas podían interferir entre sí para producir figuras de bandas luminosas y oscuras. Pues bien, las figuras que Foroughi y yo hemos visto se deben a este tipo de interferencias luminosas.

Consideremos dos rayos luminosos que iluminan una gotita suspendida en las proximidades de cada uno de los lados del rayo cartesiano que finalmente emerge formando el arco iris primario. Elijamos unos rayos que se reflejen una vez dentro de la gotita (precisamente como el rayo cartesiano) y que emerjan en la misma dirección. (Por supuesto, la dirección no será la misma que la del rayo cartesiano, pero si el observador se encuentra en el ángulo correcto, verá ambos rayos.) Cuando los rayos salen del láser están en fase, es decir, las ondas que representan están sincronizadas exactamente. No obstante, al separarse saliendo de la gotita los rayos recorren distancias diferentes y, como resultado, emergen de ella con una relación de fase que puede haber variado respecto a la original.

Supongamos que uno de los rayos haya viajado una semilongitud de onda más que el otro. (En un cálculo exacto se incluiría la variación de la longitud de onda de la luz cuando ésta se encuentra en el agua.) Cuando los rayos emerjan, el primer rayo estará rezagado respecto al segundo en una semilongitud de onda. Puesto que un observador en una posición correcta verá ambos rayos a la vez, éstos se interfieren destructivamente, lo que significa que el observador advertirá oscuridad porque los rayos (que difieren en una semilongitud de onda) están exactamente en oposición de fase.

Consideremos seguidamente la situación cuando el observador se desplaza ligeramente hacia ángulos de desvia-



Forma en que se superponen las bandas principales en una gota pequeña

ción mayores. Otra vez podrán verse dos rayos luminosos (otros distintos ahora) emergiendo de la gotita. Supongamos que la distancia viajada de más por uno de ellos sea una longitud de onda completa. Cuando dos rayos iguales se encuentran inicialmente sincronizados, un desplazamiento de uno de ellos de una longitud de onda completa lo coloca nuevamente en sincronismo. Entonces, los rayos se interfieren de una forma constructiva y el observador verá una banda luminosa, y si la luz es roja el observador verá una banda roja luminosa.

Al aumentar el ángulo de desviación las bandas luminosas y oscuras se alternan. La primera banda oscura, contigua a la luminosa principal, aparece por la interferencia de dos rayos defasados una semilongitud de onda. La segunda banda oscura se debe a otros dos rayos defasados 3/2 de longitud de onda, la tercera y cuarta bandas se deben a rayos defasados 5/2 y 7/2 de longitud de onda, respectivamente, y así sucesivamente.

Los rayos que emergen formando la banda más luminosa y ancha están en fase. La siguiente banda luminosa del espectro resulta de dos rayos que emergen defasados una longitud de onda completa. Las demás bandas luminosas del espectro resultan de otros rayos defasados una longitud de onda adicional cada vez. Si un observador se coloca frente a la figura de interferencia formada por la gotita de agua, verá siempre una banda luminosa o una oscura producida por dos rayos que emergen de la gotita. Y si mira directamente a ésta, verá dos manchas de luz en el lugar por donde emergen ambos rayos.

Si se desplaza hacia el ángulo del arco iris, verá que las dos manchas se mueven acercándose una a otra y que acaban por fundirse para constituir el rojo del arco iris.

Si se ilumina la gotita con luz blanca, cada color produce una figura de interferencia y sólo se aprecian las bandas luminosas principales de cada uno de ellos. Los primeros cálculos de la intensidad luminosa de los espectros los realizó George B. Airy en 1838. Sus resultados correspondientes a cuatro colores se ofrecen en la ilustración de la página anterior. Cada longitud de onda genera una figura de interferencia cuya banda principal se halla próxima al ángulo que predijeron las teorías de Young y Descartes para el color correspondiente del arco iris.

En los cálculos de Airy se tiene en cuenta el tamaño de la gotita, puesto que éste determina si los rayos que emergen de ella bajo un ángulo diferente del ángulo del arco iris dan luminosidad u oscuridad. El tamaño de la gotita determina también el ángulo exacto bajo el que aparece el arco iris; así, cuanto menor sea la gotita, tanto mayor resultará la discrepancia entre los ángulos de los colores y los ángulos predichos por las teorías anteriores. A medida que una gota disminuye de tamaño por evaporación, los espectros de los distintos colores cambian de dos maneras: los espectros de los arcos iris de los dos primeros órdenes se separan entre sí y las bandas luminosas se hacen más anchas. Ninguno de estos dos cambios está predicho en las teorías primitivas, en las que no describe el arco iris como una figura de interferencia.

Oprimiendo el émbolo de la jeringa, aumenta de tamaño la gotita y puede verse sin dificultad el desplazamiento resultante de las figuras. A medida que la gotita se hace mayor, las figuras de los arcos iris de los dos primeros órdenes se irán acercando. También se ve fácilmente que la anchura de cada banda brillante disminuye. Dejando que la gotita se encoja lentamente por evaporación, tienen lugar los correspondientes cambios en sentido contrario, como Foroughi observó.

La dependencia de las figuras respecto del tamaño de la gota tiene otra consecuencia. Cuando la gota es relativamente grande (de algunos milímetros de diámetro), son distinguibles las bandas brillantes principales en las figuras separadas de cada uno de los colores. Todas las bandas de cada uno de los colores son muy estrechas. Cuando en la gota penetra luz blanca el observador ve un arco iris bajo el ángulo del arco iris. Si la gota es más pequeña las bandas principales se ensanchan y se solapan y, entonces, el observador ya no puede distinguir los colores con tanta facilidad y el arco iris parece como desteñido [*véase la ilustración de esta misma página*].

Si la gotita es aún más pequeña, de un diámetro de 0,3 milímetros o menos, las bandas principales se solapan tanto que el arco iris es blanco. Y este es el arco iris blanco por el que me preguntó mi abuela. Su aparición, muy poco frecuente, sólo se da en lloviznas de gotitas suficientemente pequeñas para producir el solapamiento de las bandas principales bajo el ángulo del arco iris. Si en mi experimento se deja que la gota suspendida de la jeringa se evapore, puede verse cómo se destiñen lentamente los colores del arco iris.

Se han propuesto teorías acerca del arco iris más modernas y exactas que las de Descartes, Young y Airy. En 1908 Gustav Mie ideó un sistema para calcular las amplitudes de las ondas luminosas dispersadas por una gota.

Estos cálculos prometían resultados más exactos que las teorías anteriores, pero requieren tantas operaciones que ni siquiera los ordenadores modernos de gran velocidad pueden llevarlos a término. Para gotas de varios milímetros de diámetro las teorías antiguas daban resultados suficientemente aceptables; las dificultades aparecían al considerar gotas pequeñas. Hace tres años se expuso en esta revista un enfoque del tema más factible, conocido como teoría del momento cinético complejo del arco iris [véase "Teoría del arco iris", por H. Moisés Nussenzveig; INVESTIGACION Y CIENCIA, junio de 1977]. En la bibliografía final aportamos otras referencias acerca de la teoría del arco iris.

La luz de un arco iris natural está fuertemente polarizada en la dirección paralela a su arco. El mismo lector puede comprobarlo haciendo girar un filtro polarizador situado entre su ojo y un arco iris. Para una cierta orientación, el arco iris es muy brillante; haciendo girar el filtro 90 grados, el arco iris desaparece casi por completo.

Con la luz dispersada por una gotita de agua suspendida puede hacerse una observación similar. Si la gota se ilumina con luz procedente de un proyector, la luz incidente no está polarizada, pero puede suponerse compuesta de dos sentidos de polarización ortogonales. Uno de éstos yace en el plano definido por el proyector, el observador y la sección transversal de la gotita atravesada por el rayo luminoso. El otro sentido es perpendicular a este plano. La luz incidente polarizada en el primero de los sentidos es dispersada muy pobremente por la gotita y contribuye poco al arco iris. La luz incidente polarizada perpendicularmente al plano es la responsable fundamental del arco iris.

La luz producida por mi láser está polarizada verticalmente a causa de las ventanas de Brewster instaladas dentro de la cavidad del láser, disposición muy corriente en los láseres de gas. Por tanto, la luz está convenientemente polarizada para que atraviese una gotita suspendida por una sección transversal horizontal y cree un arco iris de colores brillantes. Al girar el tubo del láser se ve que la imagen del arco iris se hace más mortecina. Girando el láser 90 grados, el haz queda polarizado paralelamente al plano del láser, de la gotita y del observador y, entonces, el sentido de polarización de la luz que penetra en la gotita no es el adecuado y la figura del arco iris se debilita tanto que apenas es visible.

Libros

Iberia e Hispania, el firmamento visto por un aficionado y una filosofía de la biología

Domingo Fletcher, M.ª Asunción Catalá y Luis Alonso

HISTORIA DE ESPAÑA, dirigida por Manuel Tuñón de Lara. Introducción. Primeras Culturas e Hispania Romana. Editorial Labor, S.A. Barcelona, abril de 1980. 478 páginas y 26 mapas. Con este volumen inicia la editorial Labor la publicación de una HISTORIA DE ESPAÑA en diez volúmenes, en cuya redacción, bajo la dirección científica del doctor Tuñón de Lara, colaboran veintiocho prestigiosas figuras de la investigación histórica española.

En la introducción general, Tuñón de Lara expone cuáles son los propósitos al editar esta obra, explicándonos sus puntos de vista sobre lo que debe ser la historia, la que no ha de servir para pasar unas horas de lecturas más o menos gratas, sino para facilitar la comprensión de nuestro pasado "porque la historia pesa con fuerza sobre nuestra vida cotidiana", y rechaza su utilización como arma política que ha sido causa del falseamiento de los hechos "sin el menor respeto a los métodos científicos". Por ello, es propósito de editores y colaboradores "desmontar manipulaciones y desmitificar conceptos".

La necesidad de unos hitos cronológicos para mejor orientarnos en el devenir histórico fuerza la ficticia división en edades, períodos, ciclos, etcétera, que, aunque convencionales, nos ayudan a situar hombres y hechos; ante esta necesidad, Tuñón de Lara razona el porqué de la periodicidad de esta obra, escrita no sólo para los especialistas sino también para el lector con una formación cultural básica. El contenido de esta introducción abrirá al lector nuevas perspectivas sobre el moderno concepto de la historia.

La primera parte del volumen, dedicado a las Primeras Culturas, consta de introducción, doce capítulos, apéndice bibliográfico y diez mapas, con un total de 150 páginas. En aquélla, el autor, doctor Tarradell, plantea cuáles han de ser las áreas territoriales a estudiar, las cronologías y nomenclaturas, a lo que sigue un esquema del desarrollo de los estudios de prehistoria en España, haciendo mención especial del hundimiento del "mito africano" según el cual se atribuía a Africa un papel fundamental en la formación de las culturas peninsulares.

El capítulo II es un breve resumen del Paleolítico y Mesolítico, sin aportar los más recientes resultados porque, como el autor manifiesta, aunque muchos puntos de vista han sido superados, ha de esperarse un tiempo prudencial para incluir las novedades en las obras de síntesis. No obstante, creemos que se debió insistir en la trascendencia que la Cova del Parpalló (Gandía) tiene en el Paleolítico Superior europeo, con sus "puntas parpallonenses" y el más rico conjunto de losetas pintadas y grabadas conocido hasta la fecha. También al "arte rupestre levantino" pudo dársele un más amplio tratamiento, pues no hay que olvidar que es una expresión artística única, sobre la que gravitan grandes problemas, no siendo el menos importante el de su cronología, que no queda resuelto basándonos en la fauna representada.

El complejo mundo neolítico se estudia en el capítulo III. Quedan bien dibujadas las características básicas, tanto en lo estrictamente arqueológico como en las estructuras socioeconómicas, planteándose los problemas de la introducción de la agricultura, ganadería, cerámica, vivienda al aire libre, etcétera. Dos grandes etapas lo configuran, según el autor. Correspondería la primera a la llamada "cerámica cardial " y la segunda a los denominados "sepulcros de fosa". Entre los más destacados yacimientos de aquélla figuran en lugar preferente la Cova de la Sarsa (Bocairente) y la Cova de l'Or (Beniarrés, Alicante). Los "sepulcros de fosa" sólo se identifican con seguridad en Cataluña. Tal vez hubiera sido oportuno dar a conocer la debatida cuestión de una posible etapa precardial con cerámicas lisas y las nuevas conclusiones sobre un momento avanzado del neolítico caracterizado por las cerámicas con decoración incisa. De todos modos, se ha de reconocer el gran esfuerzo de síntesis que significa este capítulo, así como el siguiente, en el que se tratan los complejos problemas que presentan el Eneolítico y la Edad del Bronce; en aquél, Tarradell nos habla de la búsqueda y explotación de yacimientos metalíferos, de la defensa de los alimentos, de la inseguridad de la vida, que obliga a establecerse en lugares de fácil defensa, etcétera, destacando que el cobre y estaño juegan el "papel que el hierro y carbón han desarrollado en el siglo pasado". La variedad de las fuentes de riqueza origina los grupos con personalidad definida que se refleja en las diferentes manifestaciones culturales; así, puede distinguirse un área y época del megalitismo (de la segunda mitad del tercer milenio al año 2000 a.C.), en la que destaca el grupo de Andalucía, la que pasa, desde entonces, a ser el primer centro cultural y económico peninsular, predominio que mantendrá hasta el fin de la dominación árabe; la costa oriental, de Murcia a Barcelona, carece de construcciones megalíticas, ofreciendo el grupo valenciano una evidente personalidad, con economía agrícola frente al predominio pastoril de Cataluña. Muy acertada está su "sospecha de una cierta inflación" en los estudios del "vaso campaniforme", del que todavía quedan muchas cuestiones por resolver, a pesar de las múltiples publicaciones sobre el tema.

En la Edad del Bronce se distinguen el área del Bronce Valenciano (tierras valencianas con extensión hacia Aragón), el Argar (propio del S.E. español), las Motillas manchegas y el Bronce Meridional Portugués; sintetiza las características de los tres primeros grupos, hace referencia del cuarto sólo en el mapa III y no mencionan el llamado "Bronce Atlántico", del que hablará al tratar de las invasiones indoeuropeas llegadas a través de los Pirineos a partir del s. IX a.C. Por los pasos orientales, entran los portadores de los "campos de urnas", estableciéndose en Cataluña con esporádicas extensiones hacia el norte de la provincia de Castellón. La presencia de estos indoeuropeos tendrá repercusiones en la constitución de los pueblos ibéricos de Cataluña. Por tierras de Navarra y Lérida aparecen otros grupos que en lugar de incinerar en urnas, inhuman bajo túmulos. En la Meseta el impacto indoeuropeo es fundamental para comprender la historia posterior de aquella zona. En el N.O. no pueden señalarse claras influencias étnicas y culturales de la oleada propiamente céltica, hecho paradójico puesto que la opinión popular viene conside-

rando a Galicia como la región céltica por excelencia. A pesar de todos estos grandes movimientos de gentes motivados por la llegada indoeuropea, destaca el autor que una de las regiones que conserva "su fondo mediterráneo con mayor fuerza es la valenciana, donde precisamente veremos surgir uno de los focos más potentes y, desde luego, el más característico del iberismo", conclusión que destacamos por su trascendencia al valorar la formación y raíces de los pueblos de España.

En cuanto a las colonizaciones, fenicios y griegos hacen acto de presencia en la península a partir del 800 a.C., fecha que rebajaríamos ante la realidad arqueológica. Unas veces son pequeñas factorías costeras, otras mercados temporales o pequeños núcleos de población, siendo los minerales la atracción de unos y otros, intercambiando tejidos, joyas, objetos de adorno, cerámica, etcétera, debiendo funcionar bien este sistema de intercambio, puesto que la moneda no aparece hasta la segunda mitad del s. IV a.C. Son los fenicios los primeros que llegan a Andalucía y costa oriental ya en los s. VII/VI, no haciéndolo los griegos antes del VI. Con respecto a Tartessos, llega a la conclusión de que el tema se ha desorbitado, ya que cualquier resto material hallado en Andalucía central u occidental queda clasificado como tartésico sin matizar lo que es indígena y lo que deriva de los círculos fenicios.

El capítulo dedicado a los Iberos es el más extenso de esta primera parte de la obra. En él se dan a conocer su cronología (formación de la personalidad ibérica a partir del s. V a.C.), su área geográfica (desde el S.E. de Francia hasta Andalucía oriental) y "tribus" que la integran, organización social, poblados, necrópolis, economía, religión, arte (con la aceptación de dos etapas, la de decoración geométrica y la decoración fitomorfa, zoomorfa y figuras humanas que se desarrolla bajo la dominación romana). Todo ello está expuesto con claridad, percatándose el lector de la trascendencia que el pueblo ibérico va a tener en la posterior conformación histórica de España. Tal vez hubiera sido idóneo, para completar esta amplia visión, añadir dos apartados más: el de la moneda y el de la lengua y escritura, tan importante para identificar lo estrictamente ibérico y determinar su área geográfica frente a los pueblos limítrofes.

No es tarea fácil, dada la variedad de gentes, el estudio de los pueblos del centro y norte de la península, alejados de las corrientes civilizadoras de fenicios, griegos y cartagineses, por cuya razón mantienen sus estructuras socioeconómicas, religión y lengua, vinculadas al mundo indoeuropeo hasta su romanización. El autor ha salido airoso de esta dificultad resumiendo lo mucho escrito sobre aquellos pueblos, aproximándonos a su conocimiento.

En el proceso de romanización de Hispania, reconoce tres etapas básicas: la de la conquista por Roma del litoral mediterráneo desde los Pirineos a Andalucía, que finalizaría hacia el 200 a.C.; la conquista de los territorios cántabros, que puede considerarse terminada hacia el 25 a.C. y la tercera que es la de asimilación y adaptación de los modos de vida del mundo romano, lo que tuvo lugar a lo largo de dos siglos y con distintas cronologías, según las zonas. Destaquemos la novedad y acierto de incluir una panorámica de conjunto de la arqueología de los archipiélagos españoles, lo que no se había hecho hasta ahora en obras de esta índole. Tarradell, con su gran capacidad de síntesis (recuérdese su bello opúsculo sobre Ibiza), expone la prehistoria de las islas Canarias y las Baleares, de forma clara y asequible, a pesar de la complejidad del tema.

Bien sabemos lo difícil que es en una obra de conjunto tratar con amplitud todos los temas que se abordan. Por ello, al hacer este breve resumen, hemos creído conveniente destacar algunas cuestiones insuficientemente tratadas, no por desconocimiento del autor, sino por la forzada limitación de espacio de estas obras generales.

Terminamos la reseña de esta primera parte del libro manifestando nuestra extrañeza al comprobar que de un total de 114 títulos que se citan en la bibliografía no aparezca ningún autor valenciano tratando temas específicamente valencianos (a excepción del número 33 en que se menciona genéricamente el "Corpus Vasorum Hispanorum"), extrañeza justificada porque nos consta que el autor conoce a la perfección la bibliografía valenciana, que ha utilizado ampliamente para sus estudios, al menos en sus años de estancia en Valencia.

A partir de la página 197 se inicia la segunda parte, dedicada a la Hispania Romana. Su autor, el profesor Julio Mangas, divide su estudio en tres grandes apartados, cada uno de los cuales comprende, a su vez, tres capítulos, precedidos por una introducción, con un total de 260 páginas, nueve mapas de doble página y siete de media página. En la introducción, el autor, uno de nuestros mejores especialistas en estos temas, expone los problemas que se plantean en la utilización de los textos literarios y epigráficos y los restos arqueológicos, para el estudio de esta etapa de nuestra historia.

Después de exponer en líneas generales la conquista de Hispania, amplía la cuestión en el capítulo I, fijando la posición de ésta en la lucha entre romanos y cartagineses, cuyo punto de partida fue la toma de la ciudad de Sagunto por estos últimos. Teniendo en cuenta que, precisamente, este hecho fue la motivación de la entrada de los romanos en Hispania y de que se está estudiando su ocupación y romanización, tal vez hubiera sido conveniente que el autor se extendiera sobre las motivaciones del asedio y avatares del mismo, identificación de los límites establecidos en el "tratado del Ebro", tan debatido, y en la no menos debatida supuesta alianza romano-saguntina.

Considera el autor que la resistencia de los indígenas ha sido utilizada como símbolo de amor a la independencia y libertad, pero "afortunadamente son pocos los que siguen creyendo en estos tópicos de las virtudes de la raza" y "hoy nadie admite" que se prolongara el asedio de Numancia "ante la fiereza y el valor militar de los indígenas". La verdad es que, sin contar si son pocos o muchos los que puedan creer o no en estos "tópicos", al menos hemos de reconocer que fueron "duros de pelar" tanto en el asedio de Sagunto (nueve meses de lucha), donde las "tensiones políticas en el seno de la oligarquía romana" nada tenían que ver, como en la lucha contra Viriato, al que sólo vencieron asesinándolo, y no porque sean "virtudes" de ninguna "raza" sino simplemente porque es propio de los hombres amar la independencia y libertad, y esto no creemos que se les pueda negar a los indígenas que lucharon contra el ejército romano.

En la expedición a Galicia, Décimo Bruto asentó a las vencidas tropas de Viriato en Valentia, que el autor identifica "probablemente" en Valença do Minho; páginas más adelante vuelve a esta "probable" ubicación, pero en la página 257 la sitúa ya en la costa mediterránea, situación que mantiene en las páginas 351, 366 (mapa) y 367. Nos queda la duda de si hemos de entender que hubo dos Valentiae, lo que no parece "probable", aparte de que el emplazamiento en Valença do Minho ha sido rechazado por la crítica, que mantiene la tesis de que la del Miño se denominó, hasta 1262, Contrasta; en cambio, la mediterránea queda avalada por la epigrafía, numismática y citas clásicas, nada de lo cual puede aducirse en favor de Valença do Minho.

En el capítulo III, dedicado a las guerras civiles y conquista del territorio cántabro, el autor, que domina el tema a la perfección, nos da a conocer las rivalidades entre los bandos romanos y su repercusión en Hispania; sin embargo, es en los siguientes apartados cuando la aportación del profesor Mangas llega a sus más altas cotas. Conocedor a fondo de la administración romana y de las cuestiones sociales y económicas nos adentra en la compleja problemática de la lenta romanización de Hispania, explicando la organización político-administrativa, siguiendo con la económico-social, la minería, la agricultura, artesanado, comercio, etcétera. Paralelamente a estas transformaciones de la vida material se producen los cambios en las ideas y cultura, aunque también puede hablarse de una relativa hispanización de las formas culturales romanas de Hispania. En muchos lugares de ésta perviven las organizaciones sociales indígenas, con sus prácticas y creencias religiosas, pero en el área ibérica hubo un importante acercamiento entre la religión romana y las aquí existentes y, aunque hay santuarios ibéricos que perviven aún dentro de los primeros siglos del Imperio, son excepcionales las lápidas votivas a divinidades indígenas. Otros hechos nos permiten comprender los cambios operados en el campo cultural: en la franja costera del sudeste aún se hablaba libio-fenicio al comienzo del Imperio; en el resto de la Bética y casi todo el este peninsular, el vehículo predominante de comunicación oral es la lengua latina (aunque la ibérica no se había perdido como lo prueban los grafitos ibéricos que hemos hallado sobre terra sigillata); incluso en la Meseta había ya muchos latinoparlantes. Es decir, que antes de comenzar el período imperial, Hispania no sólo había sido sometida a un nuevo sistema político, económico y social, sino que gran parte de su población había asimilado las creencias religiosas, lengua y comportamientos sociales de los dominadores, proceso que se acelerará durante el Imperio, en cuyo estudio destaca dos etapas; en la primera se sigue, a través de dos siglos, la estructura global de comienzos del Imperio, aunque con los naturales reajustes; la segunda se inicia con la crisis del siglo III a.C. y termina con Domiciano, dando lugar a la aparición de un nuevo sistema.

Sobre la citada crisis debemos destacar la muy acertada opinión del autor para quien las correrías de francos y alemanes por tierras de Hispania causarían grandes trastornos, pero cree excesivo el que se les atribuyan todas las destrucciones acaecidas en el siglo III, suponiendo mejor que son las bandas armadas de campesinos y esclavos las causantes de tales calamidades, estudiando, con gran dominio del tema, del que es especialista, toda la problemática del papel de los esclavos y libertos en la economía y organización social del territorio hispano.

El contenido de esta parte del libro es de un nivel tan denso que cada apartado constituye por sí una sólida monografía, cuya detenida lectura nos permite comprender cómo Hispania fue acomodándose al mundo romano.

Puestos a hacer alguna observación, podríamos señalar la escasa atención prestada a la parte arqueológica. Tal vez por ello no hace referencia a la amplia red de acueductos y canales romanos de ambas márgenes del Turia, o cita únicamente la terra sigillata hispánica de Tricio cuando en el libro de Mèzquiriz se hace mención de muchos talleres de este tipo cerámico, o con respecto a la "bella cerámica" saguntina "de tradición ibérica" no se tiene en cuenta que todavía no hay acuerdo sobre a qué clase de cerámica hacía alusión Marcial, etcétera.

También podemos observar que, en el apartado de bibliografía, de los 319 títulos citados no aparece ninguno debido a investigadores valencianos ni directamente relacionados con temas valencianos, a excepción del 129 y 143, quedando sin mencionar multitud de trabajos como, por ejemplo, los de Arföldy, y cuanto se ha escrito sobre Sagunto y la fundación de Valencia.

Unas tablas cronológicas y los índices onomástico y toponímico finalizan la obra, en la que debiera incrementarse, a nuestro juicio, la parte gráfica, pues sólo mapas es poca ilustración para un libro de esta importancia y pulcra presentación.

Finalmente, para tener en cuenta en nuevas ediciones, señalamos algunos errores que, aunque intrascendentes, mejor es que no aparezcan. Por ejemplo, pág. 57, dice Superior, debe ser Inferior; 194, Barcelona, es Madrid; 266, (Valencia), es (Alicante); 265, 266 y 464, Alcazarejos, es Alcaracejos; 418 y 471, (Santa Eulalia), es (Santa Elena), etcétera. (D.F.)

GUIA DEL FIRMAMENTO, por José Luis Comellas. Ediciones Rialp, S.A. Madrid, 1979. El libro está escrito para astrónomos aficionados por un astrónomo aficionado y apasionado por la astronomía, como se desprende de su lectura ya desde la primera página. Se propone servir de ayuda para todos aquellos que sintiendo curiosidad por los temas astronómicos y quizá poseyendo manuales de astronomía, catálogos, mapas de estrellas y algún instrumento de observación, más o menos modesto, no sacan todo el provecho que pudieran de este material por falta de orientación. En el libro el lector encontrará todo lo que se ve en el firmamento y cómo se ve con un pequeño anteojo de aficionado.

Los dos primeros capítulos van dirigidos especialmente al observador novel. Empieza abordando la forma de orientarse en la Tierra; prosigue con la esfera celeste y la orientación en la misma. Se ocupa luego de las coordenadas celestes y el movimiento diurno, conceptos que quizás hubieran quedado más claros para el principiante si el capítulo hubiese sido escrito en forma más esquemática. En el segundo describe distintos tipos de instrumentos y sus condiciones ópticas. Sin despreciar la observación visual, importante auxiliar de la orientación, pasa a recomendar el uso de unos buenos prismáticos, como ayuda al observador novel, para recorridos generales por el cielo y sobre todo para la observación de la Vía Láctea. Se extiende a continuación en el estudio de anteojos y telescopios, sin tratar la teoría de los mismos, pero dando con todo detalle consejos prácticos para su manejo y para obtener buenas fotografías del cielo. Pasa después a explicar qué se entiende por abertura de un telescopio, distancia focal del mismo, campo y aumento. Acompaña al capítulo una serie de fotografías de telescopios reflectores y refractores de distintos tipos, con explicación de sus distintas partes, accesorios y montura.

Los capítulos III, IV y V, dedicados a la observación del Sol, de la Luna y los planetas, en el que incluye la de los cometas, están bien logrados; lástima que no se hayan incorporado más fotografías del Sol, mostrando las manchas solares, fáculas, granulación y oscurecimiento del borde, fenómenos todos ellos fáciles de observar por un aficionado. En el primero, tras referirse a distintos métodos de observación, pasa a explicar la orientación del disco solar y accidentes que en él se pueden apreciar, haciendo un estudio muy detallado de las manchas solares y la actividad solar. Expone a continuación la forma de observar un eclipse de Sol, valorando los fenómenos que ocurren durante el mismo y cómo fotografiarlo.

Es interesante la magnífica excursión que de mano del autor, experimentado guía, se hace por la superficie de la Luna, basada en la edad de la misma.

Aquí se presenta con todo detalle la orografía de la cara visible de la Luna, tanto que me atrevería a decir que si se leyera este capítulo a un invidente llegaría a conocer nuestro satélite como si lo tuviera en sus manos.

En el capítulo dedicado a la observación de los planetas explica, en primer lugar, cómo buscar e identificar un planeta si no se dispone de un anuario. En cuanto a los métodos de observación, aunque reconoce que es poco lo que puede hacer un aficionado, recomienda fijarse en todos los detalles que se ven en la superficie del planeta (manchas, bandas de nubes, casquetes polares, etcétera) y describirlos cuidadosamente en el cuaderno de observación o hacer un croquis o dibujo de los mismos. Para animar al lector da una secuencia de nueve dibujos de Marte correspondientes a una rotación de dicho planeta realizados por diversos astrónomos amateurs. En cuanto a la superficie de Júpiter, da con todo detalle sus accidentes (bandas y zonas y la mancha roja). Indica también cómo observar los satélites de Júpiter, aportando los elementos necesarios para ello, y los fenómenos que se pueden distinguir teniendo en cuenta su posición con respecto al planeta. En lo concerniente a Saturno llama la atención, como es natural, sobre la visión de sus anillos, uno de los fenómenos más espectaculares en la imagen de cualquier planeta y de los que más cautivan a la persona que por primera vez contempla ese planeta a través de un telescopio. Describe sus distintas posiciones hasta su desaparición aparente en alguna época. Quizá la parte más interesante de este capítulo sea la consagrada a los cometas, pero sabe a poco. El autor, que se ha empleado a fondo en el estudio de Marte y de Júpiter, no da a esta parte del libro la extensión que se merece, quizá basándose en la idea de que un aficionado poco puede hacer en este campo porque raras veces tenemos ocasión de contemplar un cometa espectacular. Sin embargo, como muy bien apunta, una buena proporción de los cometas que se descubren cada año lo son por astrónomos aficionados. Lo único que se necesita para descubrir nuevos cometas es mucha paciencia y un anteojo de gran campo y alta luminosidad.

La parte más extensa del libro está destinada a la descripción del mundo estelar, clasificando los objetos celestes más interesantes para un aficionado en: estrellas variables, estrellas dobles o múltiples, cúmulos (abiertos o globulares), nebulosas (difusas o planetarias) y galaxias. La "Guía" no incluye cuerpos o fenómenos celestes que no puedan ser observados con un modesto instrumento de aficionado. El autor, en el capítulo VI de su obra, explica cómo debe procederse para observar estrellas variables, aconsejando empezar por aquellas cuya localización en el cielo no ofrece dificultad, cómo debe procederse para observar estrellas dobles, dando alguna regla sencilla para medir la distancia angular entre sus componentes y llamando también la atención sobre el contraste de colores que podemos observar entre las dos estrellas. Hace una incursión en el terreno de los cúmulos, orientando al observador novel acostumbrado a ver los cúmulos fotografiados en los tratados de astronomía y da una clasificación lógica, muy curiosa, para uso del aficionado, en cúmulos resolubles, semirresolubles e irresolubles. En relación a la observación de nebulosas advierte que las mejores imágenes de las mismas no se obtienen por observación directa, sino mediante fotografía. De todos modos aconseja no perderse el magnífico espectáculo que proporciona la observación visual de la nebulosa de Orión. Y para vivir la emoción del infinito, según palabras del autor, el astrónomo aficionado no tiene más remedio que dedicarse a la observación de galaxias, universos como el nuestro o mucho mayores, situados a enormes distancias de nosotros. De ellas, con un refractor de 7,5 centímetros de diámetro, se pueden ver hasta 500.

A partir del capítulo VII, estudia el casquete circumpolar norte. Clasifica luego las observaciones en "campañas de primavera, de verano, de otoño y de invierno", recordando en cada una de ellas las constelaciones y objetos visibles durante los meses de las correspondientes estaciones para un observador situado en el hemisferio norte. Estas descripciones van acompañadas de mapas muy detallados de las correspondientes regiones del cielo estudiadas.

Las zonas celestes descritas llegan hasta una declinación de −40°, lo cual hace que el libro sea totalmente útil para un observador europeo, norteamericano o japonés. Para un observador situado en América Central la utilidad es de un 80 por ciento y para un observador situado en Argentina o en Australia es de algo más del 50 por ciento; es decir, más de la mitad de los objetos descritos en él se podrán observar con sus instrumentos. La razón de este límite se debe a que la obra está basada en la experiencia personal del autor, quien no se ha situado nunca, en lugares desde los que se vea el hemisferio sur de la esfera celeste por debajo de dicha declinación.

El libro está pensado para que pueda ser utilizado por lo menos hasta el año 2000. Recoge una tabla de eclipses de Sol hasta dicho año 2000 y se adjuntan, como apéndices, una lista de 174 estrellas de referencia elegidas de entre las más brillantes del cielo, para que puedan ser utilizadas como puntos de referencia en la localización de cualquier tipo de objetos, una lista de estrellas variables, otra de estrellas dobles, que es quizás el catálogo más completo que se ha hecho hasta el presente para aficionados, y una postrera de cúmulos y nebulosas. Todas ellas incluyen objetos celestes tanto boreales como australes con referencia al equinoccio de 1950,0 y de 2000,0.

El libro está, en general, muy bien logrado, y hasta podríamos definirlo como un tratado de astronomía práctica si hubiera incluido al principio un poco más de teoría, aunque hubiese sido en detrimento de la extensión de algunas descripciones, sobre todo de la parte que se refiere a las estrellas, lo que ahorraría al aficionado el tener que consultar manuales de astronomía para entender algunos de sus pasajes.

De todos modos, resulta una verdadera "Guía del firmamento", refleja los años de experiencia que como astrónomo aficionado lleva su autor y contagia al lector su entusiasmo por el estudio del cielo. A ello ayuda la magnífica colección de fotografías, sobre todo de la Luna, realizadas por astrónomos aficionados, que intercala en el texto. (M.A.C.)

FILOSOFIA DE LA BIOLOGIA, por Michael Ruse. Alianza Editorial, Madrid, 1979. ¿Es la biología una ciencia autónoma o hay que reducirla a la física y a la química? ¿Cuáles son los problemas de la biología que le interesan al filósofo? ¿Habrá que debatir todavía quién tiene razón, si el vitalista o el mecanicista? Para responder a esas cuestiones se han venido publicando desde 1970 obras como la que reseñamos, la de D. Hull (*Philosophy of Biological Science*), B. Rensch (*Biop-Biological Science*), B. Rensch (*Biophilosophy*), F. J. Ayala, T. Dobzhans-*Biology*), R. Bernier y P. Pirlot (*Organe et fonction*) y quizá la compilación más completa de M. Grene y E. Mendelsohn (*Topics in the Philosophy of Biology*), amén de la revista *History and Philosophy of the Life Sciences*. La generalidad de los autores enfocan y solucionan los problemas de acuerdo con su formación de origen (filosofía), con la excepción del Ayala que une a su

prestigio en el campo de la genética de poblaciones una sólida cultura humanística.

La biología nutre sus contenidos teóricos con presupuestos que le ofrece la física y la química. Lo biológico es en primera instancia algo físico o químico. Los ácidos nucleicos que encierran el código vital están constituidos por bases nitrogenadas, pentosa y ácido fosfórico: productos todos ellos hábiles en cualquier laboratorio. Su comportamiento está regulado por enzimas, proteínas de función catalizadora asimismo sujetas a las leyes que gobiernan las macromoléculas. La célula, el organismo y el ecosistema se mueven y rigen por principios que, en la mayoría de los casos, son de orden termodinámico. La biología será, pues, reducible a la física y a la química, se concluirá.

Pero he aquí que llegan entonces los filósofos de la biología y lo complican con el sí, pero no. "¿Podrían esperar los biólogos moleculares presentar, algún día, un análisis físico-químico completo del orden biológico? Aun cambiando sus teorías, ¿tendrán que considerar siempre que cierto orden es un dato? Personalmente opino que sí... Sin embargo, muchas áreas de la Biología, por ejemplo, la Sistemática y la Paleontológica, continúan ocupándose esencialmente de problemas puramente biológicos y pueden dar soluciones apoyadas solamente en teorías que en sí mismas son biológicas" (Ruse, página 260). Ayala, con cautela, de la que también participa nuestro autor, deja la solución para un mañana indeterminado: "La reducción de la biología a la fisicoquímica no puede llevarse a cabo en el estado actual de los conocimientos científicos. Y el que la reducción sea posible en el futuro constituye una cuestión empíricamente carente de sentido" ("Biology as an autonomous Science", en *American Scientist, Sigma XIX*, 56, p. 207-221, 1968). Opinión que mantenía diez años más tarde ("El reduccionismo en biología", en *Arbor*, noviembre de 1978, págs. 24-37).

Esa indecisión respecto a la naturaleza de la biología comporta, por fuerza, provisionalidad y eclecticismo en problemas filosóficos relacionados con ella: organismo, teleología, causalidad, etcétera. Otros temas, sin embargo, admiten un tratamiento más firme. En concreto, el de la nueva sistemática. En oposición a la taxonomía tipológica o fijista se entiende por tal el método de clasificación de los organismos por razón de su dotación génica y su adaptabilidad selectiva. Desde Linneo, y antes, los taxones se encuadraban en distintas categorías de acuerdo con su morfología externa. Práctica que hemos seguido todavía en los laboratorios de los departamentos de zoología y botánica y en los museos de historia natural.

La genética mendeliana, la teoría darwinista de la evolución y la genética molecular y de poblaciones constituyeron sobre todo los pilares en que se asentaba la nueva sistemática defendida por Mayr y Simpson. Aunque hubo antecedentes, pero sobre todo a raíz de la publicación en 1963 de la obra de Sneath y Sokal *Principles of Numerical Taxonomy,* surge la escuela fenetista, que incorporaba los métodos matemáticos a la clasificación de los organismos. Ruse los contrapone y opta salomónicamente por la vía media de darle validez a unos en ciertos casos y a otros en otras circunstancias.

Pero Ruse no supo apreciar la escuela cladística encabezada por Willi Hennig y defendida, entre legión, por S. G.Kiriakoff, L. Brundin, R. A. Crowson y S. J. Nelson. El análisis cladístico consiste en la rigurosa aplicación del principio siguiente: una especie x es un pariente más cercano a una especie y que a una especie z, si y sólo si posee al menos una especie ancestral en común con la y, antepasado que no es al propio tiempo ancestro de la especie z. Este enfoque se ha erigido en el más serio opositor del método fenetista y contra él va dirigida la segunda edición (1973) de la obra de Sneath y Sokal, publicada, pues, el mismo año de la aparición del libro que comentamos. El planteamiento de la situación que hiciera Ruse quedó ya obsoleto el día que salió a la luz pública su filosofía de la biología. Los seis primeros capítulos, sin embargo, mantienen incluso hoy plena validez en línea de principio y de rigor metodológico, aunque por ser predominantemente biológicos se resienten del tiempo en que fueron redactados. No sería justo, por tanto, achacarle omisiones como la reordenación cromosómica y los elementos genéticos transponibles a la hora de explicar el cambio genético.

El libro español sale con errores tipográficos que afean no pocas de sus páginas. En la 40, por ejemplo, el desarrollo de la ley de Hardy-Weinberg está mal transcrito en tres pasos; el ARN ribosómico aparece traducido por "ribosoma ARN (rARN)"; faltan flechas en la página 246; etcétera. Son detalles que debieran corregirse en una segunda edición, en cuya puesta al día debe estar empeñado el autor consagrado ya en la materia con una obra sobre sociobiología, publicada en inglés recientemente. (L.A.)

Bibliografía

Los lectores interesados en una mayor profundización de los temas expuestos pueden consultar los trabajos siguientes:

MICORRIZAS

THE BIOLOGY OF MICORRHIZA. J. L. Harley. Leonard Hill Press, 1972.

THE *ENDOGONACEAE* IN THE PACIFIC NORTHWEST. J. W. Gerdemann y J. M. Trappe en *Mycologia Memoir n.° 5*. The New York Botanical Garden, 1974.

ENDOMYCORRHIZAS. Editado por F. E. Sanders, B. Mosse y P. B. Tinker. Academic Press, 1975.

INTERACTIONS BETWEEN NON-PATHOGENIC SOIL MICROORGANISMS AND PLANTS. Editado por Y. Dommergues y S. V. Krupa. (Especialmente cap. 10, D. S. Hayman) Elsevier Scientific Publishing Company, 1978.

FERTILIZANTES MICROBIANOS: INTERACCIONES DE *RHIZOBIUM* Y HONGOS DE LAS MICORRIZAS V-A EN LA FORMACION Y EFICACIA DE SUS RESPECTIVAS SIMBIOSIS CON LEGUMINOSAS. C. Azcón-G. de Aguilar (dirigida por J. M. Barea). Tesis Doctorales de la Universidad de Granada, 1980.

METEORITOS BASALTICOS

A PETROGENETIC MODEL OF THE RELATIONSHIPS AMONG ACHONDRITIC METEORITES. Edward M. Stolper, Harry Y. McSween, Jr., y James F. Hays en *Geochimica et Cosmochimica Acta,* vol. 43, n.° 4, págs. 589-602; abril, 1979.

ALLAN HILLS 77005: A NEW METEORITE TYPE FOUND IN ANTARCTICA. Harry Y. McSween, Jr., Lawrence A. Taylor y Edward M. Stolper en *Science,* vol. 204, n.° 4398, págs. 1201-1203; 15 de junio de 1979.

PETROLOGY AND ORIGIN OF THE SHERGOTTITE METEORITES. Edward M. Stolper y Harry Y. McSween, Jr., en *Geochimica et Cosmochimica Acta,* vol. 43, n.° 9, págs. 1475-1498; septiembre, 1979.

CHAQUETE POR ORDENADOR

SOME NECESSARY CONDITIONS FOR A MASTER CHESS PROGRAM. Hans J. Berliner en *The Third International Joint Conference on Artificial Intelligence.* Stanford Research Institute, 1973.

ON THE CONSTRUCTION OF EVALUATION FUNCTIONS FOR LARGE DOMAINS. Hans Berliner en *The Fifth International Joint Conference on Artificial Intelligence.* Stanford Research Institute, 1979.

FISIOLOGIA DE LAS PLAQUETAS SANGUINEAS

PLATELET BEHAVIOR AND BIOCHEMISTRY. H. Holmsen, L. Salganicoff y M. H. Fukami en *Haemostasis: Biochemistry, Physiology, and Pathology,* dirigido por D. Ogston y B. Bennett. John Wiley & Sons, 1977.

PLATELETS, THROMBOSIS AND ATHEROSCLEROSIS. J. F. Mustard, S. Moore, M. A. Packham y R. L. Kinlough-Rathbone en *Progress in Biochemical Pharmacology,* vol. 13, págs. 312-325; 1977.

TEORIAS GAUGE DE LAS FUERZAS ENTRE PARTICULAS ELEMENTALES

CONSERVATION OF ISOTOPIC SPIN AND ISOTOPIC GAUGE INVARIANCE. C. N. Yang y R. L. Mills en *Physical Review,* vol. 96, n.° 1, págs. 191-195; 1 de octubre de 1954.

UNIFIED THEORIES OF ELEMENTARY-PARTICLE INTERACTION. Steven Weinberg en *Scientific American,* vol. 231, n.° 1, págs. 50-59; julio, 1974.

QUARKS WITH COLOR AND FLAVOR. Sheldon Lee Glashow in *Scientific American,* vol. 233, n.° 4, págs. 38-50; octubre, 1975.

SUBSISTENCIA EN EL NORTE DE ESPAÑA

THE FOOD CRISIS IN PREHISTORY: OVERPOPULATION AND THE ORIGINS OF AGRICULTURE. Mark Nathan Cohen. Yale University Press, 1977.

LATE AND POST-PLEISTOCENE INDUSTRIES AND FAUNA FROM THE CAVE SITE OF LA RIERA. G. A. Clark y L. Richards en *Views of the Past: Essays in Old World Prehistory and Paleoanthropology.* Ninth International Congress of Anthropological and Ethnological Sciences, dirigido por Leslie G. Freeman. Mouton, 1978.

LA HIPOTESIS DEL GAS DE LAS PROFUNDIDADES TERRESTRES

DUPLEX ORIGIN OF PETROLEUM. Sir Robert Robinson en *Nature,* vol. 199, n.° 4889, págs. 113-114; 13 de julio de 1963.

BRONTIDES: NATURAL EXPLOSIVE NOISES. Thomas Gold y Steven Soter en *Science*, vol. 204, n.° 4391, págs. 371-375; 27 de abril de 1979.

SELECCION SEXUAL EN *HYLOBITTACUS APICALIS*

SEXUAL SELECTION AND NUPTIAL FEEDING BEHAVIOR IN *BITTACUS APICALIS* (INSECTA: MECOPTERA). Randy Thornhill en *American Naturalist,* vol. 110, n.° 974, págs. 529-548; julio-agosto, 1976.

INSECT BEHAVIOR. Robert W. Matthews y Janice R. Matthews. John Wiley & Sons, 1978.

MALE AND FEMALE SEXUAL SELECTION AND THE EVOLUTION OF MATTING STRATEGIES IN INSECTS. Randy Thornhill en *Sexual Selection and Reproductive Competition in Insects,* dirigido por Murray S. Blum y Nancy A. Blum. Academic Press, 1979.

JUEGOS MATEMATICOS

FINITE GROUPS. Daniel Gorenstein. Harper and Row, 1968.

THE SEARCH FOR FINITE SIMPLE GROUPS. Joseph A. Gallian en *Mathematics Magazine,* vol. 49, n.° 4, págs. 163-180; septiembre, 1976.

GROUPS AND SYMMETRY. Jonathan L. Alperin en *Mathematics Today,* dirigido por Lynn Arthur Steen. Springer-Verlag, 1978.

TALLER Y LABORATORIO

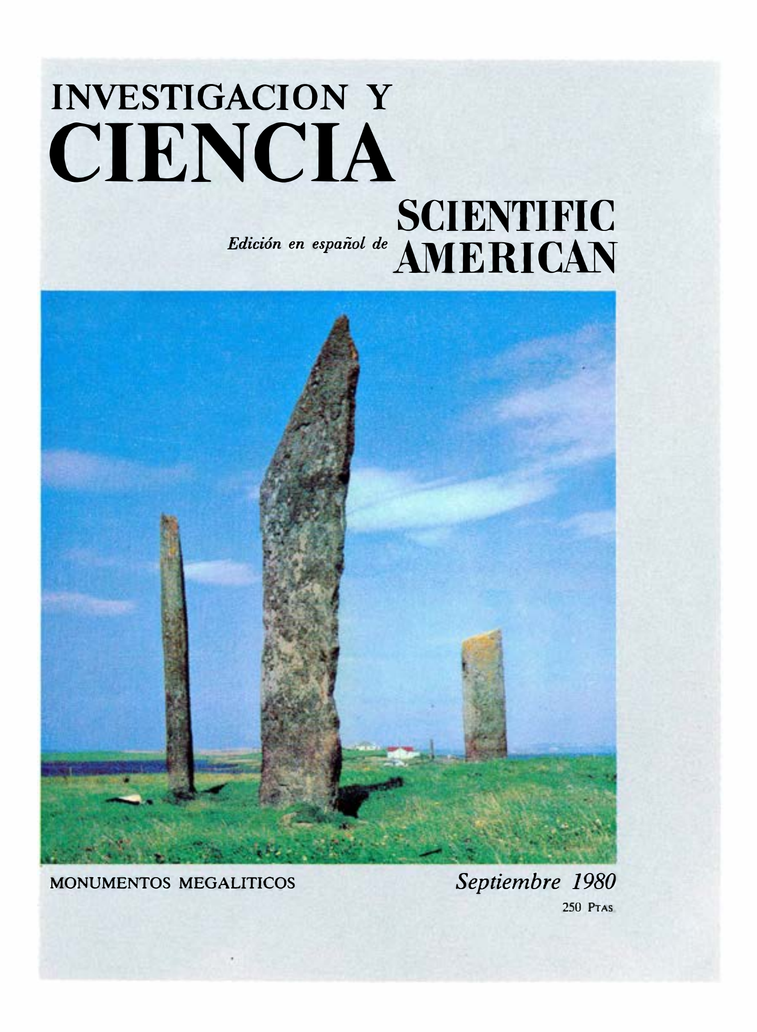
THE RAINBOW: FROM MYTH TO MATHEMATICS. Carl B. Boyer. Thomas Yoseloff, 1959.

MULTIPLE RAINBOWS FROM SINGLE DROPS OF WATER AND OTHER LIQUIDS. Jearl D. Walker en *American Journal of Physics,* vol. 44, n.° 5, págs. 421-433; mayo, 1976.

ANGULAR SCATTERING AND RAINBOW FORMATION IN PENDANT DROPS. Kenneth Sassen en *Journal of the Optical Society of America,* vol. 69, n.° 8, págs. 1083-1089; agosto, 1979.

COMPLEX ANGULAR MOMENTUM THEORY OF THE RAINBOW AND THE GLORY. H. M. Nussenzveig en *Journal of the Optical Society of America,* vol. 69, n.° 8, págs. 1068-1079; agosto, 1979.

Seguiremos explorando los campos del conocimiento



ADN SUPERENROLLADO, por William R. Bauer, F.H.C. Crick y James H. White
En muchas formas del ADN la doble hélice forma a su vez una hélice. Un sencillo modelo matemático describe y analiza este superenrollamiento, que tiene importantes consecuencias biológicas.

PROPULSION A HELICE, por Eugene Larrabee
Las hélices de los barcos y los aviones parecen bastante diferentes, pero ambas están diseñadas según la teoría circulatoria del ascenso. Esta teoría es la que tradicionalmente se desarrolló para explicar el ascenso de una superficie aerodinámica.

ASTRONOMIA DE LINEAS DE RAYOS GAMMA, por Marvin Leventhal y Crawford J. MacCallum
Los procesos astrofísicos emiten radiación electromagnética en bandas muy estrechas a energías superiores incluso a las de los rayos X. Las líneas espectrales resultantes se pueden ahora detectar mediante instrumentos montados en globos o satélites.

MONUMENTOS MEGALITICOS, por Glyn Daniel
Estos conjuntos de enormes piedras, de los que Stonehenge es uno, se encuentran a miles en Europa. Su edad fue mucho tiempo incierta, pero ahora han sido fechados en época neolítica.

FISIOLOGIA DEL KOALA, por Robert Degabriele
Este dócil marsupial come principalmente hojas de eucalipto (tóxicas para otros animales), bebe rara vez y no utiliza refugio. Estudios recientes indican cómo se las arregla para sobrevivir en su nicho característico.

LUPUS ERITEMATOSO SISTEMICO, por David Koffler
El LES, una enfermedad reumática con múltiples síntomas, es una alteración del sistema inmunitario. El paciente produce autoanticuerpos que se unen a antígenos como el ADN formando complejos inmunitarios que pueden lesionar los tejidos.

LA TRADUCCION DE LOS LENGUAJES DE COMPUTADOR, por Manuel Alfonseca y M.ª Luisa Tavera
Los traductores de los lenguajes de alto nivel utilizados en los computadores suelen ser programas grandes y complejos, y conviene hacerlos independientes de la máquina en que han de ser ejecutados, para evitar la duplicación de esfuerzos.

CONVECCION, por Manuel García Velarde y Christiane Normand
Sólo será posible entender las corrientes ascendentes y descendentes en un fluido calentado, cuando se describan las complejas relaciones que existen entre temperatura, viscosidad, tensión superficial y demás propiedades del fluido.

